Another Look at Success Probability in Linear Cryptanalysis

Subhabrata Samajder
R. C. Bose Center for Cryptology and Security
Indian Statistical Institute
203, B.T.Road, Kolkata, India - 700108.
subhabrata.samajder@gmail.com

Palash Sarkar
Applied Statistics Unit
Indian Statistical Institute
203, B.T.Road, Kolkata, India - 700108.
palash@isical.ac.in

Abstract

This work studies the success probability of linear cryptanalysis. Complete expressions for the success probability are obtained using two different approaches, namely the order statistics and the hypothesis testing based approaches. We argue that the hypothesis testing based approach is theoretically more sound and does not require a number of assumptions and approximations which are inherent in the order statistics based approach. For analysing success probability, a unifying framework of general key randomisation hypotheses is introduced. The previously used standard key randomisation hypotheses and the adjusted wrong key randomisation hypothesis can be seen to special cases of the general framework. Derivations of expressions for the success probability are carried out under both the settings of the plaintexts being sampled with and without replacements. Finally, the complete picture of the dependence of the success probability on the data complexity is derived and it is argued that in most practical scenarios, the data complexity will be a monotone increasing function of the data complexity. We believe that compared to the extant literature, our work provides a deeper and more thorough understanding of the success probability of linear cryptanalysis.
Keywords: linear cryptanalysis, success probability, data complexity.

1 Introduction

A block cipher is a fundamental cryptographic primitive. Such a primitive injectively maps an n-bit plaintext under the influence of a secret key to an n-bit ciphertext. The strength of a block cipher is assessed by determining its resistance to the known standard attacks.

Linear cryptanalysis [20] is a fundamental method of attacking a block cipher. To apply linear cryptanalysis, it is required to first obtain an approximate linear relation between the input and the output of a block cipher. Obtaining such a relation for a well designed cipher is a non-trivial task and requires a great deal of ingenuity along with a very careful examination of the internal structure of the mapping which defines the target block cipher. The present work does not address this aspect of linear cryptanalysis and it will be assumed that a linear relation is available.

The goal of (linear) cryptanalysis of a block cipher is to recover a portion of the secret key in time less than that required by a brute force algorithm to try out all possible keys. The portion of the key which is proposed to be recovered is called the target sub-key. An attack with such a goal is called a key recovery attack. A weaker goal is to be able to distinguish the output of the block cipher from that of a uniform random permutation and such attacks are called distinguishing attacks. In this work, we will concentrate only on key recovery attacks.

To apply linear cryptanalysis, it is required to obtain some data corresponding to the secret key. Such data consists of plaintext-ciphertext pairs (P_i, C_i), $i = 1, \ldots, N$, where C_i is obtained by encrypting P_i using the secret key. The plaintexts are chosen randomly. Typically, they are considered to be chosen under uniform random sampling with or without replacements.

Any method of determing the secret key from this data is statistical in nature. The output of the attack is a set of candidate values for the target sub-key. The attack is successful with some probability P_S if the correct value of the target sub-key is in the set of candidate values. The size of the set of candidate values is also an important parameter. An attack is said to have a-bit advantage if the size of the set of candidate values is a fraction 2^{-a} of the number of possible values of the target sub-key [26].

The goal of a statistical analysis of an attack is to be able to obtain a relation between the three fundamental parameters N, P_S and a. In this work, we concentrate on obtaining P_S as a function of N and a and closely examine the behaviour of P_S as a function of N.

A linear approximation of a block cipher holds with certain probability. Broadly speaking, a key recovery attack proceeds by testing each value of the target sub-key against the linear approximation with respect to the available data. For the correct choice κ^* of the target sub-key, the linear approximation holds with some probabilty p_{κ^*} while for an incorrect choice $\kappa \neq \kappa^*$ of the target sub-key, the linear approximation holds with some other probability p_{κ,κ^*}. The basis of the attack is a difference in p_{κ^*} and p_{κ,κ^*}. The detailed examination of the internal structure of the block cipher leads to an estimate of p_{κ^*}, while p_{κ,κ^*} is obtained from an analysis of the behaviour of a random mapping.

To perform a statistical analysis, it is required to hypothesise the values of p_{κ^*} and p_{κ,κ^*}. The hypothesis on p_{κ^*} is called the right key randomisation hypothesis, while the hypothesis on p_{κ,κ^*} is called the wrong key randomisation hypothesis. Until a few years ago, it was typical to hypothesis that p_{κ^*} is a constant $p \neq 1/2$ while $p_{\kappa,\kappa^*} = 1/2$. These are called the standard right and wrong key randomisation hypothesis respectively.

The adjusted wrong key randomisation hypothesis was introduced by Bogdanov and Tischhauser in [8]. Based on a previous work by Daemen and Rijmen [9], it was hypothesised that p_{κ,κ^*} itself is a random variable following the normal distribution $\mathcal{N}(1/2, 2^{-n-2})$. A later work by Ashur, Beyne and Rijmen [1] also used the adjusted wrong key randomisation hypothesis. The difference in [8] and [1] is in the manner in which the plaintexts $P_1, \ldots, P_N$ were assumed to be chosen – sampling with replacement was considered in [8] while sampling without replacement was considered in [1]. Both the works [8, 1] observed a non-monotonic dependence of the success probability on N and provided possible explanations for this phenomenon.

The actual statistical methodology used in [8, 1] is based on an earlier work by Selçuk [26] which was based on the use of order statistics. This, in turn, was a formalisation of a ranking methodology used in the original work of Matsui [20] which introduced linear cryptanalysis.

Our Contributions

We perform a complete and generalised analysis of success probability in linear cryptanalysis.

Complete expression for the success probability: The expression for success probability obtained by Selçuk [26] is incomplete. Suppose $\mathcal{S}$ is the event that an attack is successful. In [26], Selçuk works considers a sub-event of $\mathcal{S}$ to be $\mathcal{S}$ leading to an incomplete expression for the success probability. The later works [8, 1] follow Selçuk's approach and hence also obtain incomplete expressions for the success probability. In the present work, we obtain the complete expression for the success probability.

Statistical methodology: The expression for the success probability can be derived in two different ways. The first method is based on an order statistics approach while the second method uses statistical hypothesis testing. We derive expressions for the success probability using both these methods. The expressions for the success probability obtained using the two different approaches are slightly different. They turn out to be equal if certain assumptions and approximations used by Selçuk in [26] are applied to the expression obtained from the order statistics based approach. Some theoretical limitations of the order statistics approach was pointed out in [24]. In the present work, we identify two other implicit independence assumptions that need to be made to apply this approach. In contrast, the hypothesis testing based analysis does not suffer from the theoretical

limitations and nor are any assumptions or approximations required. So, from a theoretical point of view, the hypothesis testing based approach is more satisfying. Consequently, we take the expression obtained from the hypothesis testing based approach to be the correct expression for the success probability.

General key randomisation hypotheses: Following the formalisation of the adjusted wrong key randomisation hypothesis, we introduce the general key randomisation hypotheses. The general right key randomisation hypothesis models p_{κ^*} as a random variable following $\mathcal{N}(p, s_0^2)$ and the general wrong key randomisation hypothesis models p_{κ,κ^*} as a random variable following $\mathcal{N}(1/2, s_1^2)$. The expression for success probabilty is obtained in terms of p, s_0 and s_1. Letting $s_0, s_1 \downarrow 0$, we obtain the standard key randomisation hypothesis while setting $s_1 = 2^{-n-2}$ gives the adjusted wrong key randomisation hypothesis. As a result, from the general expression for the success probability, we are able to obtain particular expressions for the success probability under standard key randomisation hypotheses and under the adjusted wrong key randomisation hypothesis. Further, our method of analysis covers both the cases of sampling with and without replacements allowing us to obtain expressions for the success probability under both these sampling strategies. We note that the standard right key randomisation hypothesis has been extended to the adjusted right key randomisation hypothesis in the context of multiple linear cryptanalysis [6].

Analysis of non-monotonicity of the success probability: We perform a general analysis of the dependence of the success probability on N. We show that if the Fisher information about the mean p in the random variable p_{κ^*} is not more than the Fisher information about the mean $1/2$ in the random variable p_{κ,κ^*}, then the success probability is a monotone increasing function of N. On the other hand, depending on the relative values of s_0, s_1 and p, we obtain cases where the success probability indeed decreases with increasing N. By using appropriate values of s_0 and s_1, we particularise the analysis to the case of adjusted wrong key randomisation hypothesis for both sampling with and without replacements. The complete picture of the dependence of the success probability on N is worked out in both these cases. Under a relatively mild assumption on the magnitude of p, it can be shown that the success probability is again a monotonic increasing function of N. For the adjusted wrong key randomisation hypothesis, the previous analyses [8, 1] of the dependence of success probability on N did not reveal the complete picture that this described in this work.

Previous and Related Work

Linear cryptanalysis was first proposed by Matsui in [20]. Junod [15] gave a detailed analysis of Matsui's ranking method [20, 21]. This work introduced the notion of ordered statistics in linear cryptanalysis. The idea was further developed by Selçuk in [26], where he used a well known asymptotic result from the theory of ordered statistic to arrive at an expression for the success probability. Building on a work by Daemen and Rijmen [9], a paper by Bogdanov and Tischhauser [7] introduced the adjusted wrong key randomisation hypothesis. The work [7] considered the plaintexts to be sampled with replacement. A later work by Ashur, Beyne and Rijmen [1] analysed success probability under adjusted wrong key randomisation hypothesis in the setting where the plaintexts are sampled without replacements.

Analysis of attacks using multiple linear approximations has been reported in the literature [21, 18, 4, 17, 2, 16, 3, 11, 22, 13, 6, 14, 24, 25, 23]. Since this paper is concerned only with the basic setting of a single linear approximation, we do not discuss the various aspects which arise in the context of multiple linear approximations.

2 Linear Cryptanalysis: Background and Statistical Model

Let $E : \{0,1\}^k \times \{0,1\}^n \mapsto \{0,1\}^n$ denote a block cipher such that for each $K \in \{0,1\}^k$, $E_K(\cdot) \stackrel{\Delta}{=} E(K, \cdot)$ is a bijection from the set $\{0,1\}^n$ to itself. Here K is called the secret key. The n-bit input to the block cipher is

called the plaintext and n-bit output of the block cipher is called the ciphertext.

Block ciphers are generally constructed by composing round functions where each round function is parametrised by a round key. The round functions are also bijections of $\{0,1\}^n$ to itself. The round keys are produced by applying an expansion function, called the key scheduling algorithm, to the secret key K. Denote the round keys by $k^{(0)}, k^{(1)}, \ldots$ and the round functions by $R^{(0)}_{k^{(0)}}, R^{(1)}_{k^{(1)}}, \ldots$. For $i \geq 1$, let $K^{(i)}$ denote the concatenation of the first i round keys, i.e., $K^{(i)} = k^{(0)} \,||\, \cdots \,||\, k^{(i-1)}$ and $E^{(i)}_{K^{(i)}}$ denote the composition of the first i round functions, i.e., $E^{(1)}_{K^{(1)}} = R^{(0)}_{k^{(0)}}$ and for $i \geq 2$, $E^{(i)}_{K^{(i)}} = R^{(i-1)}_{k^{(i-1)}} \circ \cdots \circ R^{(0)}_{k^{(0)}} = R^{(i-1)}_{k^{(i-1)}} \circ E^{(i-1)}_{k^{(i-1)}}$.

A block cipher may have many rounds and for the purposes of estimating the strength of a block cipher, a cryptanalytic attempt may target only some of these rounds. Such an attack is called a reduced round cryptanalysis. Suppose an attack targets the first $r+1$ rounds where the block cipher may possibly have more than $r+1$ rounds. For a plaintext P, we denote by C the output after $r+1$ rounds, i.e., $C = E^{(r+1)}_{K^{(r+1)}}(P)$, and by B the output after r rounds, i.e., $B = E^{(r)}_{K^{(r)}}(P)$ and $C = R^{(r)}_{k^{(r)}}(B)$.

Linear approximation: Any block cipher cryptanalysis starts off with a detailed analysis of the structure of the block cipher. This results in one or more relations between the plaintext P, the input to the last round B and possibly the expanded key $K^{(r)}$. In case of linear cryptanalysis a linear relation of the following form is obtained.

$$\langle \Gamma_P, P \rangle \oplus \langle \Gamma_B, B \rangle \quad = \quad \langle \Gamma_K, K^{(r)} \rangle. \tag{1}$$

where $\Gamma_P, \Gamma_B \in \{0,1\}^n$ and $\Gamma_{K^{(r)}} \in \{0,1\}^{nr}$ denote the plaintext mask, the mask to the input of the last round and the key mask.

A relation of the form given by (1) is called a linear approximation of the block cipher. Such a linear approximation usually holds with some probability which is taken over the random choices of the plaintext P. Obtaining such a linear approximation and the corresponding probability is a non-trivial task and requires a lot of ingenuity and experience. This forms the basis on which the statistical analysis of block ciphers is built.

Define

$$L \stackrel{\Delta}{=} \langle \Gamma_P, P \rangle \oplus \langle \Gamma_B, B \rangle. \tag{2}$$

Inner key bit: Let

$$z = \langle \Gamma_K, K^{(r)} \rangle.$$

Note that for a fixed but unknown key $K^{(r)}$, z is a single unknown bit. Since the key mask Γ_K is known, the bit z is determined only by the unknown but fixed $K^{(r)}$. Hence, there is no randomness in either of $K^{(r)}$ or z. The bit z is called the inner key bit.

Target sub-key: A linear relation of the form (1) usually involves only a subset of the bits of B. In order to obtain these bits from the ciphertext C it is required to partially decrypt C by one round. This involves a subset of the bits of the last round key $k^{(r)}$. We call this subset of bits of the last round key to be the target sub-key.

The ciphertext C is obtained by encrypting P using a key K. By κ^* we denote the value of the target sub-key corresponding to the key K. We are interested in a key recovery attack where the goal is to find κ^*.

Let the size of the target sub-key be m. These m bits are sufficient to partially decrypt C by one round and obtain the bits of B involved in the linear approximation. There are 2^m possible choices of the target sub-key out of which only one is correct. The purpose of the attack is to identify the correct value.

Probability and bias of a linear approximation: Let P be a plaintext chosen uniformly at random from $\{0,1\}^n$; C be the corresponding ciphertext; and B be the result of partially decrypting C with a choice κ of the target sub-key. The random variable B depends on the choice κ that is used to partially invert C. Further, C depends on the correct value κ^* of the target sub-key and hence so does B. So, the random variable L defined in (2) depends on κ and κ^* and we write L_{κ,κ^*} to emphasise this dependence. For $\kappa = \kappa^*$, we will simply write L_{κ^*}. Define

$$p_{\kappa,\kappa^*} = \Pr[L_{\kappa,\kappa^*} = 1],\ \kappa \neq \kappa^*; \qquad p_{\kappa^*} = \Pr[L_{\kappa^*} = 1]; \tag{3}$$

$$\epsilon_{\kappa,\kappa^*} = p_{\kappa,\kappa^*} - 1/2; \qquad \epsilon_{\kappa^*} = p_{\kappa^*} - 1/2. \tag{4}$$

Here $\epsilon_{\kappa,\kappa^*}$ and ϵ_{κ^*} are the biases corresponding to incorrect and correct choices of the target sub-key respectively. The secret key K is a fixed quantity and so the randomness arises solely from the uniform random choice of P.

Statistical model of the attack: Let $P_1, \ldots, P_N$, with $N \leq 2^n$, be chosen randomly following some distribution from the set $\{0,1\}^n$ of all possible plaintexts. It is assumed that the adversary possesses the N plaintext-ciphertext pairs $(P_j, C_j); j = 1, 2, \ldots, N$ where $C_j = E_K(P_j)$ for some fixed key K. Using the linear approximation and the N plaintext-ciphertext pairs, the adversary has to find κ^* in time faster than a brute force search on all possible keys of the block cipher.

For each choice κ of the target sub-key it is possible for the attacker to partially decrypt each C_j by one round to obtain $B_{\kappa,j}; j = 1, 2, \ldots, N$. Note that $B_{\kappa,j}$ depends on κ even though C_j may not do so. Clearly, if $\kappa = \kappa^*$, then the C_j's depend on κ, while if $\kappa \neq \kappa^*$, C_j has no relation to κ.

For $\kappa \in \{0, 1, \ldots, 2^m - 1\}$, $z \in \{0,1\}$, $j = 1, \ldots, N$, define

$$L_{\kappa,j} = \langle \Gamma_P, P_j \rangle \oplus \langle \Gamma_B, B_{\kappa,j} \rangle; \tag{5}$$

$$X_{\kappa,z,j} = L_{\kappa,j} \oplus z; \tag{6}$$

$$X_{\kappa,z} = X_{\kappa,z,1} + \cdots + X_{\kappa,z,N}. \tag{7}$$

Note that $X_{\kappa,z,j} \oplus X_{\kappa,1\oplus z,j} = 1$ and so $X_{\kappa,0} + X_{\kappa,1} = N$.

$X_{\kappa,z,j}$ is determined by the pair (P_j, C_j), the choice κ of the target sub-key and the choice z of the inner key bit. Since C_j depends upon K and hence upon κ^*, $X_{\kappa,z,j}$ also depends upon κ^* through C_j. The randomness in $X_{\kappa,z,j}$ arises from the randomness in P_j and also possibly from the previous choices $P_1, \ldots, P_{j-1}$. $X_{\kappa,z,j}$ is binary valued and the probability $\Pr[X_{\kappa,z,j} = 1]$ potentially depends upon the following quantities:

z : the choice of the inner key bit;
p_{κ^*} or p_{κ,κ^*} : the probabilities of linear approximation as given in (3).
j : the index determining the pair (P_j, C_j).

This models a general scenario which captures a possible dependence on the index j. The dependence on j will be determined by the joint distribution of the plaintexts $P_1, \ldots, P_N$. In the case that $P_1, \ldots, P_N$ are independent and uniformly distributed, $\Pr[X_{\kappa,z,j} = 1]$ does not depend on j. On the other hand, suppose that $P_1, \ldots, P_N$ are sampled without replacement. In such a scenario, $\Pr[X_{\kappa,z,j} = 1]$ does depend on j.

Test statistic: For each choice κ of the target sub-key and each choice z of the inner key bit, let $T_{\kappa,z} \equiv T(X_{\kappa,z,1}, \ldots, X_{\kappa,z,N})$ denote a test statistic. Then $T_{\kappa,z}$ is a random variable whose randomness arises from the randomness of $P_1, \ldots, P_N$. Define

$$T_{\kappa,z} = |W_{\kappa,z}| \quad \text{where} \quad W_{\kappa,z} = \frac{X_{\kappa,z}}{N} - \frac{1}{2}.$$

Then

$$T_{\kappa,1} = |W_{\kappa,1}| = \left|\frac{X_{\kappa,1}}{N} - \frac{1}{2}\right| = \left|\frac{N - X_{\kappa,0}}{N} - \frac{1}{2}\right| = \left|\frac{1}{2} - \frac{X_{\kappa,0}}{N}\right| = |-W_{\kappa,0}| = T_{\kappa,0}.$$

So, the test statistic $T_{\kappa,z}$ does not depend on the value of z and it is sufficient to consider $z = 0$.
Remark: To simplify notation, we will write $X_{\kappa,j}$ and X_κ instead of $X_{\kappa,0,j}$ and $X_{\kappa,0}$ respectively; W_κ and T_κ instead of $W_{\kappa,0}$ and $T_{\kappa,0}$ respectively.

Using this notation, the test statistic T_κ is defined in the following manner.

$$T_\kappa = |W_\kappa| \quad \text{where} \quad W_\kappa = \frac{X_\kappa}{N} - \frac{1}{2} = \frac{X_{\kappa,1} + \cdots + X_{\kappa,N}}{N} - \frac{1}{2}. \tag{8}$$

This test statistic was considered by Matsui [20].

There are 2^m choices of the target sub-key and so there are 2^m random variables T_κ. The distribution of T_κ depends on whether κ is correct or incorrect. To perform a statistical analysis of an attack, it is required to obtain the distribution of T_κ under both correct and incorrect choices of κ. Later we consider this issue in more details.

Success probability: An attack will produce a set (or a list) of candidate values of the target sub-key. The attack is considered successful if the correct value of the target sub-key κ^* is in the output set. The probability of this event is called the success probability of the attack.

Advantage: An attack is said to have advantage a if the size of the set of candidate values of the target sub-key is equal to 2^{m-a}. In other words, a fraction 2^{-a} portion of the possible 2^m values of the target sub-key is produced by the attack.

Data complexity: The number N of plaintext-ciphertext pairs required for an attack is called the data complexity of the attack. Clearly, N depends on the success probability P_S and the advantage a. One of the goals of a statistical analysis is to be able to obtain a closed form relation between N, P_S and a.

Notation on normal distributions: By $\mathcal{N}(\mu, \sigma^2)$ we will denote the normal distribution with mean μ and variance σ^2. The density function of $\mathcal{N}(\mu, \sigma^2)$ will be denoted by $\mathfrak{f}(x; \mu, \sigma^2)$. The density function of the standard normal will be denoted by $\phi(x)$ while the distribution function of the standard normal will be denoted by $\Phi(x)$.

3 Success Probability in Linear Cryptanalysis

As given in (8), the test statistic is $T_\kappa = |W_\kappa|$ where $W_\kappa = (X_{\kappa,1} + \cdots + X_{\kappa,N})/N - 1/2$. To obtain the success probability of the attack it is required to obtain the distributions of T_κ for the two scenarios when $\kappa = \kappa^*$ and when $\kappa \neq \kappa^*$. This is obtained from the distributions of W_{κ^*} and W_κ for $\kappa \neq \kappa^*$. Suppose, the following holds.

$$W_{\kappa^*} \sim \mathcal{N}(\mu_0, \sigma_0^2), \quad \mu_0 \neq 0; \qquad W_\kappa \sim \mathcal{N}(0, \sigma_1^2), \quad \kappa \neq \kappa^*. \tag{9}$$

We now consider the derivation of the success probability of linear cryptanalysis in terms of μ_0, σ_0 and σ_1 using both the order statistics based analysis and the hypothesis testing based analysis. Later, we will see how to obtain μ_0, σ_0 and σ_1. In particular, we will see that σ_0 and σ_1 depend on N whereas μ_0 is a constant.

3.1 Order Statistics Based Analysis

This approach is based on a ranking methodology used originally by Matsui [20] and later formalised by Selçuk [26]. The idea is the following. There are 2^m random variables T_κ corresponding to the 2^m possible values of the target sub-key. Suppose the variables are denoted as $T_0, \ldots, T_{2^m-1}$ and assume that $T_0 = |W_0|$ corresponds to the choice of the correct target sub-key κ^*, where W_0 follows the distribution of W_{κ^*} which is $\mathcal{N}(\mu_0, \sigma_0^2)$. Let $T_{(1)}, \ldots, T_{(2^m-1)}$ be the order statistics of $T_1, \ldots, T_{2^m-1}$, i.e., $T_{(1)}, \ldots, T_{(2^m-1)}$ is the ascending order sort of $T_1, \ldots, T_{2^m-1}$. So, the event corresponding to a successful attack with a-bit advantage is $T_0 > T_{(2^m q)}$ where $q = 1 - 2^{-a}$.

Using a well known result on order statistics, the distribution of $T_{(2^m q)}$ can be assumed to approximately follow $\mathcal{N}(\mu_q, \sigma_q^2)$ where $\mu_q = \sigma_1 \Phi^{-1}(1 - 2^{-a-1})$ and $\sigma_q = \frac{\sigma_1}{2\phi(\Phi^{-1}(1-2^{-a-1}))} 2^{-(m+a)/2}$ (see Appendix A.1). Using this result, P_S can be approximated in the following manner.

$$
\begin{aligned}
P_S &= \Pr[T_0 > T_{(2^m q)}] = \Pr[|W_0| > T_{(2^m q)}] \\
&= \Pr[W_0 > T_{(2^m q)}] + \Pr[W_0 < -T_{(2^m q)}] \qquad (10) \\
&= \Pr[W_0 - T_{(2^m q)} > 0] + \Pr[W_0 + T_{(2^m q)} < 0] \\
&\approx 1 - \Phi\left(\frac{-(\mu_0 - \mu_q)}{\sqrt{\sigma_0^2 + \sigma_q^2}}\right) + \Phi\left(\frac{-(\mu_0 + \mu_q)}{\sqrt{\sigma_0^2 + \sigma_q^2}}\right) \\
&= \Phi\left(\frac{\mu_0 - \sigma_1 \Phi^{-1}(1 - 2^{-a-1})}{\sqrt{\sigma_0^2 + \sigma_q^2}}\right) + \Phi\left(\frac{-(\mu_0 + \sigma_1 \Phi^{-1}(1 - 2^{-a-1}))}{\sqrt{\sigma_0^2 + \sigma_q^2}}\right) \\
&= \Phi\left(\frac{|\mu_0| - \sigma_1 \Phi^{-1}(1 - 2^{-a-1})}{\sqrt{\sigma_0^2 + \sigma_q^2}}\right) + \Phi\left(\frac{-|\mu_0| - \sigma_1 \Phi^{-1}(1 - 2^{-a-1})}{\sqrt{\sigma_0^2 + \sigma_q^2}}\right). \qquad (11)
\end{aligned}
$$

Some criticisms: The order statistics based approach is crucially dependent on the normal approximation of the distribution of the order statistics. In the statistics literature, this result appears in an asymptotic form. Using the well known Berry-Esséen theorem, a concrete upper bound on the error in such approximation was obtained in [24]. A key observation is that the order statistics result is applied to 2^m random variables and for the result to be applied even in an asymptotic context, it is necessary that 2^m is sufficiently large. A close analysis of the hypothesis of the theorem and the error bound in the concrete setting showed the following issues. We refer to [24] for details.

- m **must be large:** This condition arises from a convergence requirement on one of the quantities in the theorem showing the result on order statistics. For the error in such convergence to be around 10^{-3}, it is required that m should be at least around 20 bits. So, if the size of the target sub-key is small, then the applicability of the order statistics based analysis is not clear.

- $m - a$ **must be large:** This condition arises from the requirement that the error in the normal approximation is small. If the error is to be around 10^{-3}, then $m - a$ should be at least around 20 bits. Recall that a is the advantage of the attack. So, for attacks with high advantage, the applicability of the order statistics based analysis is not clear.

Independence assumptions: Two assumptions are required for the analysis to be meaningful and these were implicitly used by Selçuk in [26].

1. The approximation of the distribution of the order statistic $T_{(2^m q)}$ by normal is a key step in the order statistics based approach. As mentioned above, this follows from a standard result in mathematical statistics. The hypothesis of this result requires the random variables $T_1, T_2, \ldots, T_{2^m-1}$ to be *independent* and identically distributed. It indeed holds that $T_1, T_2, \ldots, T_{2^m-1}$ are identically distributed. However, the randomness of all of these random variables arise from the randomness of $P_1, \ldots, P_N$ and so these random variables are certainly not independent. So, the independence of these random variables is a heuristic assumption.

2. Considering W_0 and $T_{(2^m q)}$ to follow normal distributions, it is assumed that $W_0 - T_{(2^m q)}$ (and $W_0 + T_{(2^m q)}$) also follows a normal distribution. A sufficient condition for $W_0 - T_{(2^m q)}$ to follow a normal distribution is that W_0 and $T_{(2^m q)}$ are independent. If W_0 and $T_{(2^m q)}$ are not independent, then it is not necessarily true that $W_0 - T_{(2^m q)}$ follows a normal distribution even if W_0 and $T_{(2^m q)}$ follow normal distributions. So, in assuming $W_0 - T_{(2^m q)}$ to follow a normal distribution, it is implicitly assumed that W_0 and $T_{(2^m q)}$ are independent. Since the randomness of both W_0 and $T_{(2^m q)}$ arise from the randomness in $P_1, \ldots, P_N$, they are clearly not independent. As a result, the assumption that $W_0 - T_{(2^m q)}$ follows a normal distribution is also a heuristic assumption.

In short, the above two assumptions can be summarised as assuming that the test statistics corresponding to different choices of the sub-key are independent. We note that such assumptions are sometimes made in the context of cryptanalysis though it is a bit surprising that the above assumptions do not seem to have been explicitly mentioned in the literature.

In later works on multiple linear and multiple differential cryptanalysis, the order statistics based analysis has been used in a number of papers [8, 13, 5, 6]. The above mentioned issues, i.e., both m and $m - a$ have to be large; and the assumption that the test statistics for different choices of the sub-key are independent, apply to all such works.

3.2 Hypothesis Testing Based Analysis

Statistical hypothesis testing for analysing block cipher cryptanalysis was carried out in [2] in the context of distinguishing attacks. For analysing key recovery attacks on block ciphers, hypothesis testing based approach was used in [24] as a method for overcoming some of the theoretical limitations of the order statistics based analysis. Subsequently, hypothesis testing based approach for analysing key recovery attacks in the context of key dependent assumptions was performed in [6].

The idea of the hypothesis testing based approach is simple and intuitive. For each choice κ of the target sub-key, let H_0 be the null hypothesis that κ is correct and H_1 be the alternative hypothesis that κ is incorrect. The test statistic $T_\kappa = |W_\kappa|$ is used to test H_0 against H_1 where the distributions of W_κ are as in (9) for both $\kappa = \kappa^*$ and $\kappa \neq \kappa^*$. The following hypothesis test is considered.

$$\left.\begin{array}{l} H_0 : \kappa \text{ is correct; versus } H_1 : \kappa \text{ is incorrect.} \\ \text{Decision rule } (\mu_0 > 0)\text{: Reject } H_0 \text{ if } T_\kappa \leq t. \\ \text{Decision rule } (\mu_0 < 0)\text{: Reject } H_0 \text{ if } T_\kappa \leq t. \end{array}\right\} \tag{12}$$

Here t is a threshold whose exact value is determined depending on the desired success probability and advantage. Such a hypothesis test gives rise to two kinds of errors: H_0 is rejected when it holds which is called the Type-1 error; and H_0 is accepted when it does not hold which is called the Type-2 error. If a Type-1 error occurs, then $\kappa = \kappa^*$ is the correct value of the target sub-key but, the test rejects it and so the attack fails to recover the correct value. So, the attack is successful if and only if Type-1 error does not occur. So, the success probability $P_S = 1 - \Pr[\text{Type-1 error}]$.

On the other hand, for every Type-2 error, an incorrect value of κ gets labelled as a candidate key. So, the number of times that Type-2 errors occurs is the size of the list of candidate keys.

Theorem 1. *Let $\kappa^* \in \{0,1\}^m$. For $\kappa \in \{0,1\}^m$, let $T_\kappa = |W_\kappa|$ be 2^m random variables, where $W_{\kappa^*} \sim \mathcal{N}(\mu_0, \sigma_0^2)$, $\mu_0 \neq 0$ and $W_\kappa \sim \mathcal{N}(0, \sigma_1^2)$ for $\kappa \neq \kappa^*$. Suppose the hypothesis test given in (12) is applied to T_κ for all $\kappa \in \{0,1\}^m$. Let $P_S = 1 - \Pr[\text{Type-1 error}]$. Then*

$$P_S = \Phi\left(\frac{|\mu_0| - \sigma_1\gamma}{\sigma_0}\right) + \Phi\left(\frac{-|\mu_0| - \sigma_1\gamma}{\sigma_0}\right) \tag{13}$$

where $\gamma = \Phi^{-1}\left(1 - \frac{2^{m-a-1}}{2^m-1}\right)$ and the expected number of times that Type-2 errors occurs is 2^{m-a}.

Proof. First assume $\mu_0 > 0$. Let $\alpha = \Pr[\text{Type-1 error}]$ and $\beta = \Pr[\text{Type-2 error}]$ and so $P_S = 1 - \alpha$. For each $\kappa \neq \kappa^*$, let Z_κ be a binary valued random variable which takes the value 1 if and only if a Type-2 error occurs for κ. So, $\Pr[Z_\kappa = 1] = \beta$. The size of the list of candidate keys returned by the test is $\sum_{\kappa \neq \kappa^*} Z_\kappa$ and so the expected size of the list of candidate keys is

$$E\left[\sum_{\kappa \neq \kappa^*} Z_\kappa\right] = \sum_{\kappa \neq \kappa^*} E\left[Z_\kappa\right] = \sum_{\kappa \neq \kappa^*} \Pr[Z_\kappa = 1] = (2^m - 1)\beta. \tag{14}$$

The expected number of times that Type-2 errors occurs is 2^{m-a}. So,

$$\beta = \frac{2^{m-a}}{2^m - 1}. \tag{15}$$

The Type-1 and Type-2 error probabilities are calculated as follows.

$$\begin{aligned}
\alpha &= \Pr[\text{Type-1 error}] \\
&= \Pr[T_\kappa \leq t | H_0 \text{ holds}] \\
&= \Pr[T_{\kappa^*} \leq t] \\
&= \Pr[|W_{\kappa^*}| \leq t] \\
&= \Pr[-t \leq W_{\kappa^*} \leq t] \qquad (16) \\
&= \Pr\left[\frac{-t - \mu_0}{\sigma_0} \leq \frac{W_{\kappa^*} - \mu_0}{\sigma_0} \leq \frac{t - \mu_0}{\sigma_0}\right] \\
&= \Phi\left(\frac{t - \mu_0}{\sigma_0}\right) - \Phi\left(\frac{-t - \mu_0}{\sigma_0}\right); \qquad (17) \\
\beta &= \Pr[\text{Type-2 error}] \\
&= \Pr[T_\kappa > t | H_1 \text{ holds}] \\
&= \Pr[|W_\kappa| > t | H_1 \text{ holds}] \\
&= \Pr[W_\kappa > t | H_1 \text{ holds}] + \Pr[W_\kappa < -t | H_1 \text{ holds}] \\
&= \Pr\left[\frac{W_\kappa}{\sigma_1} > \frac{t}{\sigma_1} | H_1 \text{ holds}\right] + \Pr\left[\frac{W_\kappa}{\sigma_1} < \frac{-t}{\sigma_1} | H_1 \text{ holds}\right] \\
&= 1 - \Phi\left(\frac{t}{\sigma_1}\right) + \Phi\left(\frac{-t}{\sigma_1}\right) \\
&= 2(1 - \Phi(t/\sigma_1)). \qquad (18)
\end{aligned}$$

Using $\beta = 2^{m-a}/(2^m - 1)$ in (18), we obtain

$$t = \sigma_1\gamma \quad \text{where} \quad \gamma = \Phi^{-1}\left(1 - \frac{2^{m-a-1}}{2^m - 1}\right). \tag{19}$$

Substituting t in (17) and noting that $P_S = 1 - \alpha$, we obtain

$$P_S = \Phi\left(\frac{\mu_0 - \sigma_1\gamma}{\sigma_0}\right) + \Phi\left(\frac{-(\mu_0 + \sigma_1\gamma)}{\sigma_0}\right) = \Phi\left(\frac{|\mu_0| - \sigma_1\gamma}{\sigma_0}\right) + \Phi\left(\frac{-|\mu_0| - \sigma_1\gamma}{\sigma_0}\right).$$

If $\mu_0 < 0$, then an analysis similar to the above shows that the resulting expression for the success probability is still given by (13). □

Remarks:

1. We have $\gamma = \Phi^{-1}\left(1 - 2^{m-a-1}/(2^m - 1)\right) \geq 0$ if and only if $1 - 2^{m-a-1}/(2^m - 1) \geq 1/2$ if and only if $a \geq \lg(2^m/(2^m - 1))$, where lg is logarithm to base two. We will be interested in attacks where the advantage a is at least $\lg(2^m/(2^m - 1))$ so that γ can be assumed to be non-negative.
2. The computation in (14) does not require the Z_κ's or the T_κ's to be independent.
3. The theoretical limitations of the order statistics based analysis (namely, m and $m - a$ are large and the heuristic assumption that the T_κ's are independent) are not present in the hypothesis testing based analysis.
4. Comparing (13) to (11), we find that the two expressions are equal under the following two assumptions:
   (a) $2^m/(2^m - 1) \approx 1$: this holds for moderate values of m, but, is not valid for small values of m.
   (b) $\sigma_0 \gg \sigma_q$: this assumption was used in [26] and we provide more details later.

In the rest of the work, we will use (13) as the expression for the success probability.

4 General Key Randomisation Hypotheses

Recall the definitions of p_{κ,κ^*} and p_{κ^*} from (3). The corresponding biases are $\epsilon_{\kappa,\kappa^*}$ and ϵ_{κ^*}. For obtaining the distributions of W_{κ^*} and W_κ, $\kappa \neq \kappa^*$, it is required to hypothesise the behaviour of p_{κ^*} and p_{κ,κ^*} respectively. The two standard key randomisation hypotheses are the following.

Standard right key randomisation hypothesis: $p_{\kappa^*} = p$, for some constant p for every choice of κ^*.
Standard wrong key randomisation hypothesis: $p_{\kappa,\kappa^*} = 1/2$ for every choice of κ^* and $\kappa \neq \kappa^*$.

The standard wrong key randomisation hypothesis was formally considered in [12], though it was used in earlier works. Modification of this hypothesis has been been considered in the literature. Based on an earlier work [9] on the distribution of correlations for a uniform random permutation, the standard wrong key randomisation hypothesis was relaxed in [8]. Under the standard wrong key randomisation hypothesis, the bias $\epsilon_{\kappa,\kappa^*} = 0$. In [8], it was suggested that instead of assuming $\epsilon_{\kappa,\kappa^*}$ to be 0, $\epsilon_{\kappa,\kappa^*}$ should be assumed to follow a normal distribution with expectation 0 and variance 2^{-n-2}. This is stated more formally as follows.

Adjusted wrong key randomisation hypothesis:

$$\text{For } \kappa \neq \kappa^*,\ \epsilon_{\kappa,\kappa^*} \sim \mathcal{N}\left(0, 2^{-n-2}\right), \text{ or, equivalently } p_{\kappa,\kappa^*} \sim \mathcal{N}\left(1/2, 2^{-n-2}\right).$$

Remarks:

1. In this hypothesis, there is no explicit dependence of the bias on either κ or κ^*.

2. From (4), $\epsilon_{\kappa,\kappa^*}$ should take values in $[-1/2, 1/2]$. If $\epsilon_{\kappa,\kappa^*}$ is assigned a value which is outside the range $[-1/2, 1/2]$, then p_{κ,κ^*} takes a value outside the range $[0, 1]$. Since p_{κ,κ^*} is a probability, this is meaningless. On the other hand, a random variable following a normal distribution can take any real value. So, the above hypothesis may lead to $\epsilon_{\kappa,\kappa^*}$ taking a value outside the range $[-1/2, 1/2]$ which is not meaningful. The reason why such a situation arises is that in [9], a discrete distribution has been approximated by a normal distribution without adjusting for the possibility that the values may fall outside the meaningful range. From a theoretical point of view, assuming $\epsilon_{\kappa,\kappa^*}$ to follow a normal distribution cannot be formally justified. Hence, the adjusted wrong key randomisation hypothesis must necessarily be considered to be a *heuristic* assumption.

3. The variance 2^{-n-2} is an exponentially decreasing function of n and by Chebyshev's inequality $\Pr[|p_{\kappa,\kappa^*} - 1/2| > 1/2] \leq 4 \cdot 2^{-n-2} = 2^{-n}$. In other words, p_{κ,κ^*} takes values outside $[0, 1]$ with exponentially low probability.

4. The formal statement of the adjusted wrong key randomisation hypothesis appears as Hypothesis 2 in [8] and is $|\epsilon_{\kappa,\kappa^*}| \sim \mathcal{N}\left(1/2, 2^{-n-2}\right)$, i.e., the condition in Hypothesis 2 of [8] is on the absolute value of $\epsilon_{\kappa,\kappa^*}$ rather than on $\epsilon_{\kappa,\kappa^*}$. Since the absolute value is by definition a non-negative quantity, it is not meaningful to model its distribution using normal. In fact, the proof of Lemma 5.9 in the thesis [27] makes use of the hypothesis without the absolute value, i.e., it uses the hypothesis as stated above. Further, the later work [1] also uses the hypothesis without the absolute value. So, in this work we will use the hypothesis as stated above and without the absolute sign.

While the adjusted wrong key randomisation hypothesis was used in [8] and later in [1] both of these works used the standard right key randomisation hypothesis. Modification of the right key randomisation hypothesis was considered in [6] in the context of multiple/multi-dimensional linear cryptanalysis. Based on the formulation in [6] and the adjusted wrong key randomisation hypothesis, it is possible to formulate an adjusted right key randomisation hypothesis. Motivated by this consideration we formulate the following general key randomisation hypotheses for both the right and the wrong key.

General right key randomisation hypothesis:

$$p_{\kappa^*} \sim \mathcal{N}\left(p, s_0^2\right) \text{ where } p \text{ is a fixed value and } s_0^2 \leq 2^{-n}; \text{ let } \epsilon = p - 1/2.$$

General wrong key randomisation hypothesis:

$$\text{For } \kappa \neq \kappa^*,\ p_{\kappa,\kappa^*} \sim \mathcal{N}\left(1/2, s_1^2\right) \text{ where } s_1^2 \leq 2^{-n}.$$

We note the following.

1. As $s_0 \downarrow 0$, the random variable p_{κ^*} becomes degenerate and takes the value of the constant p. In this case, the general right key randomisation hypothesis becomes the standard right key randomisation hypothesis.

2. As $s_1 \downarrow 0$, the random variable p_{κ,κ^*} becomes degenerate and takes the value $1/2$. In this case, the general wrong key randomisation hypothesis becomes the standard wrong key randomisation hypothesis.

3. For $s_1^2 = 2^{-n-2}$, the general wrong key randomisation hypothesis becomes the adjusted wrong key randomisation hypothesis.

So, the general key randomisation hypotheses covers both the standard (right and wrong) key randomisation hypotheses and also the adjusted wrong key randomisation hypothesis. In view of this, we perform the statistical analysis of success probability in terms of the general key randomisation hypotheses and later deduce the special

cases of the standard and the adjusted key randomisation hypotheses.

Remark: The issues discussed in Points 1 to 3 as part of the remarks after the adjusted wrong key randomisation hypothesis also hold for both the general right and the general wrong key randomisation hypotheses. In particular, we note that the requirements $s_0^2 \leq 2^{-n}$ and $s_1^2 \leq 2^{-n}$ have been imposed so that using Chebyshev's inequality, we obtain $\Pr[|p_{\kappa^*} - 1/2| > 1/2] \leq 4s_0^2 \leq 2^{-n+2}$ and $\Pr[|p_{\kappa,\kappa^*} - 1/2| > 1/2] \leq 4s_1^2 \leq 2^{-n+2}$ respectively. In other words, the requirements $s_0^2 \leq 2^{-n}$ and $s_1^2 \leq 2^{-n}$ ensure that the probabilities of p_{κ^*} and p_{κ,κ^*} taking values outside the range $[0, 1]$ is exponentially small.

5 Analysis of Success Probability

Given the behaviour of p_κ and p_{κ,κ^*} modelled by the two general key randomisation hypotheses, the main task is to obtain normal approximations of the distributions of W_{κ^*} and W_κ as given by (9). This will provide the values of μ_0, σ_0 and σ_1. Plugging in these values into the expression given by (13) provides the corresponding expression for the success probability. The distributions of W_{κ^*} and W_κ depend on whether $P_1, \ldots, P_N$ are chosen with or without replacement. We separately consider both these cases.

In the general key randomisation hypotheses, we have $s_0^2, s_1^2 \leq 2^{-n}$. Let $\theta_0^2 = s_0^2 2^{n/2} \leq 2^{-n/2}$. By Chebyshev's inequality,

$$\Pr[|p_{\kappa^*} - p| > \theta_0] \quad \leq \quad s_0^2/\theta_0^2 = 2^{-n/2}. \tag{20}$$

So, with exponentially low probability, p_{κ^*} takes values outside the range $[p - \theta_0, p + \theta_0]$. For $\mathfrak{p} \in [p - \theta_0, p + \theta_0]$ and $\theta = \mathfrak{p} - 1/2$, we have $\epsilon - \theta_0 \leq \theta \leq \epsilon + \theta_0$ and so

$$\mathfrak{p}(1 - \mathfrak{p}) \quad = \quad 1/4 - \theta^2 \geq 1/4 - (\epsilon + \theta_0)^2 \approx 1/4 \tag{21}$$

under the assumption that $(\epsilon + \theta_0)^2$ is negligible.

Similarly, let $\vartheta_1^2 = s_1^2 2^{n/2} \leq 2^{-n/2}$ and as above, we have by Chebyshev's inequality

$$\Pr[|p_{\kappa,\kappa^*} - 1/2| > \vartheta_1] \quad \leq \quad s_1^2/\vartheta_1^2 = 2^{-n/2}. \tag{22}$$

Further, let $\vartheta = \mathfrak{p} - 1/2$ so that for $\mathfrak{p} \in [1/2 - \vartheta_1, 1/2 + \vartheta_1]$,

$$\mathfrak{p}(1 - \mathfrak{p}) \quad = \quad 1/4 - \vartheta^2 \geq 1/4 - \vartheta_1^2 = 1/4 - s_1^2 2^{n/2} \geq 1/4 - 2^{-n/2} \approx 1/4 \tag{23}$$

under the assumption that $2^{-n/2}$ is negligible.

5.1 Distributions of W_{κ^*} and $W_\kappa, \kappa \neq \kappa^*$ under Uniform Random Sampling with Replacement

In this case, $P_1, \ldots, P_N$ are chosen under uniform random sampling with replacement so that $P_1, \ldots, P_N$ are assumed to be independent and uniformly distributed over $\{0, 1\}^n$.

First consider W_{κ^*} whose distribution is determined from the distribution of p_{κ^*}. Recall that $X_{\kappa^*} = X_{\kappa^*,1} + \cdots + X_{\kappa^*,N}$. Since $P_1, \ldots, P_N$ are independent, the random variables $X_{\kappa^*,1}, \ldots, X_{\kappa^*,N}$ are also independent. Under the general right key randomisation assumption, p_{κ^*} is modelled as a random variable following $\mathcal{N}(p, s_0^2)$

and so the density function of p_{κ^*} is $\mathfrak{f}(\mathfrak{p}; p, s_0^2)$. The distribution function of X_{κ^*} is approximated as follows:

$$\begin{aligned}
&\Pr[X_{\kappa^*} \leq x] \\
&= \sum_{\mathfrak{k} \leq x} \Pr[X_{\kappa^*} = \mathfrak{k}] \\
&\approx \sum_{\mathfrak{k} \leq x} \int_{-\infty}^{\infty} \binom{N}{\mathfrak{k}} \mathfrak{p}^{\mathfrak{k}} (1-\mathfrak{p})^{N-\mathfrak{k}} \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p} \\
&= \int_{-\infty}^{\infty} \left(\sum_{\mathfrak{k} \leq x} \binom{N}{\mathfrak{k}} \mathfrak{p}^{\mathfrak{k}} (1-\mathfrak{p})^{N-\mathfrak{k}} \right) \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p}. \qquad (24)
\end{aligned}$$

The sum within the integral is the distribution function of the binomial distribution and can be approximated by $\mathcal{N}(N\mathfrak{p}, N\mathfrak{p}(1-\mathfrak{p}))$. In this approximation, the variance of the normal also depends on $\mathfrak{p}$ which makes it difficult to proceed with further analysis. Using (21), it is possible to approximate $\mathfrak{p}(1-\mathfrak{p}))$ as $1/4$. This approximation, however, is valid only for $\mathfrak{p} \in [p-\theta_0, p+\theta_0]$ and under the assumption that $(\epsilon+\theta_0)^2$ is negligible. In particular, the approximation is not valid for values of $\mathfrak{p}$ close to 0 or 1. The probability that $\mathfrak{p}$ is not in $[p-\theta_0, p+\theta_0]$ is exponentially small as shown in (20). So, we break up the integral in (24) in a manner such that the approximation $\mathfrak{p}(1-\mathfrak{p})) \approx 1/4$ can be made in the range $p-\theta_0$ to $p+\theta_0$ and it is possible to show that the contribution to (24) for $\mathfrak{p}$ outside this range is negligible.

$$\begin{aligned}
&\Pr[X_{\kappa^*} \leq x] \\
&= \int_{p-\theta_0}^{p+\theta_0} \left(\sum_{\mathfrak{k} \leq x} \binom{N}{\mathfrak{k}} \mathfrak{p}^{\mathfrak{k}} (1-\mathfrak{p})^{N-\mathfrak{k}} \right) \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p} \\
&\quad + \int_{-\infty}^{p-\theta_0} \left(\sum_{\mathfrak{k} \leq x} \binom{N}{\mathfrak{k}} \mathfrak{p}^{\mathfrak{k}} (1-\mathfrak{p})^{N-\mathfrak{k}} \right) \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p} + \int_{p+\theta_0}^{\infty} \left(\sum_{\mathfrak{k} \leq x} \binom{N}{\mathfrak{k}} \mathfrak{p}^{\mathfrak{k}} (1-\mathfrak{p})^{N-\mathfrak{k}} \right) \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p} \qquad (25) \\
&\leq \int_{p-\theta_0}^{p+\theta_0} \left(\sum_{\mathfrak{k} \leq x} \binom{N}{\mathfrak{k}} \mathfrak{p}^{\mathfrak{k}} (1-\mathfrak{p})^{N-\mathfrak{k}} \right) \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p} + \int_{-\infty}^{p-\theta_0} \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p} + \int_{p+\theta_0}^{\infty} \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p} \\
&= \int_{p-\theta_0}^{p+\theta_0} \left(\sum_{\mathfrak{k} \leq x} \binom{N}{\mathfrak{k}} \mathfrak{p}^{\mathfrak{k}} (1-\mathfrak{p})^{N-\mathfrak{k}} \right) \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p} + \Pr[|p_{\kappa^*} - p| > \theta_0] \\
&\leq \int_{p-\theta_0}^{p+\theta_0} \left(\sum_{\mathfrak{k} \leq x} \binom{N}{\mathfrak{k}} \mathfrak{p}^{\mathfrak{k}} (1-\mathfrak{p})^{N-\mathfrak{k}} \right) \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p} + 2^{-n/2} \quad \text{(from (20))} \\
&\approx \int_{p-\theta_0}^{p+\theta_0} \left(\sum_{\mathfrak{k} \leq x} \binom{N}{\mathfrak{k}} \mathfrak{p}^{\mathfrak{k}} (1-\mathfrak{p})^{N-\mathfrak{k}} \right) \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p}. \qquad (26)
\end{aligned}$$

The sum inside the integral is approximated by the distribution function of $\mathcal{N}(N\mathfrak{p}, N\mathfrak{p}(1-\mathfrak{p}))$. The range of the integration over $\mathfrak{p}$ is from $p-\theta_0$ to $p+\theta_0$. Using (21), it follows that for $\mathfrak{p} \in [p-\theta_0, p+\theta_0]$ the normal distribution $\mathcal{N}(N\mathfrak{p}, N\mathfrak{p}(1-\mathfrak{p}))$ can be approximated as $\mathcal{N}(N\mathfrak{p}, N/4)$ (i.e., $\mathfrak{p}(1-\mathfrak{p}) \approx 1/4$) under the assumption that $(\epsilon+\theta_0)^2$ is negligible. Note that the above analysis has been done to ensure that the range of $\mathfrak{p}$ is such that

this approximation is meaningful.

$$\begin{aligned}
\Pr[X_{\kappa^*} \le x] &\approx \int_{p-\theta_0}^{p+\theta_0} \left(\int_{-\infty}^{x} \mathfrak{f}(x; N\mathfrak{p}, N/4) dx \right) \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p}. \\
&\le \int_{-\infty}^{\infty} \left(\int_{-\infty}^{x} \mathfrak{f}(x; N\mathfrak{p}, N/4) dx \right) \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p}. \qquad (27) \\
&= \int_{-\infty}^{x} \int_{-\infty}^{\infty} \left(\mathfrak{f}(x; N\mathfrak{p}, N/4) \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p} \right) \, dx \qquad (28) \\
&= \int_{-\infty}^{x} \mathfrak{f}(x; Np, s_0^2 N^2 + N/4) \, dx. \qquad (29)
\end{aligned}$$

The last equality follows from Proposition 1 in Section A.2. Comparing (24) and (27), it may appear that a roundabout route has been taken to essentially replace the sum inside the integral by a normal approximation. On the other hand, without taking this route, we do not see how to justify that the variance of this normal approximation is approximately $N/4$.

From (29), the distribution of X_{κ^*} is approximately $\mathcal{N}(Np, s_0^2 N^2 + N/4)$. Consequently, the distribution of $W_{\kappa^*} = X_{\kappa^*}/N - 1/2$ is approximately given as follows:

$$W_{\kappa^*} \sim \mathcal{N}\left(\epsilon, s_0^2 + \frac{1}{4N} \right). \qquad (30)$$

For W_κ with $\kappa \neq \kappa^*$, we need to consider the general wrong key randomisation hypothesis where p_{κ,κ^*} is modelled as a random variable following $\mathcal{N}(1/2, s_1^2)$. A similar analysis as above is carried out where instead of (20) and (21), the relations (22) and (23) respectively are used. In particular, for $\mathfrak{p} \in [1/2 - \vartheta_1, 1/2 + \vartheta_1]$, it is required to approximate $\mathcal{N}(N\mathfrak{p}, N\mathfrak{p}(1-\mathfrak{p}))$ by $\mathcal{N}(N/2, N/4)$, i.e., $\mathfrak{p}(1-\mathfrak{p}) \approx 1/4$. The validity of this approximation for $\mathfrak{p} \in [1/2 - \vartheta_1, 1/2 + \vartheta_1]$ follows from (23) where $s_1^2 2^{n/2}$ is considered to be negligible. Again, we note that the approximation $\mathfrak{p}(1-\mathfrak{p}) \approx 1/4$ is not valid for values of $\mathfrak{p}$ near to 0 or 1. The analysis yields the following approximation:

$$W_\kappa \sim \mathcal{N}\left(0, s_1^2 + \frac{1}{4N} \right), \quad \kappa \neq \kappa^*. \qquad (31)$$

Remark: For the adjusted wrong key randomisation hypothesis, i.e., with $s_1^2 = 2^{-n-2}$, in [8] the distribution of W_κ for $\kappa \neq \kappa^*$ was stated without proof to be $\mathcal{N}\left(0, \frac{1}{2^{n+2}} + \frac{1}{4N}\right)$. Lemma 5.9 in the thesis [27] also stated this result and as proof mentioned $\mathcal{N}(0, \frac{1}{2^{n+2}}) + \mathcal{N}(0, \frac{1}{4N}) = \mathcal{N}(0, \frac{1}{2^{n+2}} + \frac{1}{4N})$. This refers to the fact that the sum of two independent normal distributed random variables is also normal distributed. While this fact is well known, it is not relevant to the present analysis.

5.2 Distributions of W_{κ^*} and $W_\kappa, \kappa \neq \kappa^*$ under Uniform Random Sampling without Replacement

In this scenario, the plaintexts $P_1, \ldots, P_N$ are chosen according to uniform random sampling without replacement. As a result, $P_1, \ldots, P_N$ are no longer independent and correspondingly neither are $X_{\kappa,1}, \ldots, X_{\kappa,N}$. So, the analysis in the case for sampling with replacement needs to be modified.

We first consider the distribution of W_{κ^*} in the scenario where p_{κ^*} is a random variable. A fraction p_{κ^*} of the 2^n possible plaintexts P satisfies the condition $\langle \Gamma_P, P \rangle \oplus \langle \Gamma_B, B \rangle = 1$. Let us say that a plaintext P is 'red' if the condition $\langle \Gamma_P, P \rangle \oplus \langle \Gamma_B, B \rangle = 1$ holds for P; otherwise, we say that P is 'white'. So there are $\lfloor p_{\kappa^*} 2^n \rfloor$ red plaintexts in $\{0,1\}^n$ and the other plaintexts are white. For $\mathfrak{k} \in \{0, \ldots, N\}$, the event $X_{\kappa^*} = \mathfrak{k}$ is the event of

picking $\mathfrak{k}$ red plaintexts in N trials from an urn containing 2^n plaintexts out of which $\lfloor p_{\kappa^*} 2^n \rfloor$ are red and the rest are white. So,

$$\Pr[X_{\kappa^*} = \mathfrak{k}] = \frac{\binom{\lfloor p_{\kappa^*} 2^n \rfloor}{\mathfrak{k}} \binom{2^n - \lfloor p_{\kappa^*} 2^n \rfloor}{N-\mathfrak{k}}}{\binom{2^n}{N}}. \tag{32}$$

Under the general right key randomisation hypothesis it is assumed that p_{κ^*} follows $\mathcal{N}(p, s_0^2)$ so that the density function of p_{κ^*} is taken to be $\mathfrak{f}(\mathfrak{p}; p, s_0^2)$. Then

$$\begin{aligned}
\Pr[X_{\kappa^*} \leq x] &= \sum_{\mathfrak{k} \leq x} \Pr[X_{\kappa} = \mathfrak{k}] \\
&\approx \sum_{\mathfrak{k} \leq x} \int_{-\infty}^{\infty} \frac{\binom{\lfloor \mathfrak{p} 2^n \rfloor}{\mathfrak{k}} \binom{2^n - \lfloor \mathfrak{p} 2^n \rfloor}{N-\mathfrak{k}}}{\binom{2^n}{N}} \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p} \\
&= \int_{-\infty}^{\infty} \left(\sum_{\mathfrak{k} \leq x} \frac{\binom{\lfloor \mathfrak{p} 2^n \rfloor}{\mathfrak{k}} \binom{2^n - \lfloor \mathfrak{p} 2^n \rfloor}{N-\mathfrak{k}}}{\binom{2^n}{N}} \right) \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p}.
\end{aligned}$$

An analysis along the lines of (25) to (26) using (20) shows that

$$\Pr[X_{\kappa^*} \leq x] \approx \int_{p-\theta_0}^{p+\theta_0} \left(\sum_{\mathfrak{k} \leq x} \frac{\binom{\lfloor \mathfrak{p} 2^n \rfloor}{\mathfrak{k}} \binom{2^n - \lfloor \mathfrak{p} 2^n \rfloor}{N-\mathfrak{k}}}{\binom{2^n}{N}} \right) \mathfrak{f}(\mathfrak{p}; p, s_0^2) d\mathfrak{p}.$$

The sum within the integral can be seen to be the distribution function of the hypergeometric distribution $\mathsf{Hypergeometric}(N, 2^n, \lfloor \mathfrak{p} 2^n \rfloor)$. If $N \ll 2^n$, then the hypergeometric distribution approximately follows $\mathsf{Bin}(N, \mathfrak{p})$; on the other hand, if $N/2^n = \mathfrak{t} \in (0, 1)$, then the hypergeometric distribution approximately follows $\mathcal{N}(\mathfrak{p}N, N(1-\mathfrak{t})\mathfrak{p}(1-\mathfrak{p}))$ (see Appendix A.3) which using $\mathfrak{t} = N/2^n$ is equal to $\mathcal{N}(\mathfrak{p}N, N(1 - N/2^n)\mathfrak{p}(1-\mathfrak{p}))$.

For $\mathfrak{p} \in [p - \theta_0, p + \theta_0]$, from (21) the normal distribution $\mathcal{N}(\mathfrak{p}N, N(1 - N/2^n)\mathfrak{p}(1-\mathfrak{p}))$ is approximated as $\mathcal{N}(N\mathfrak{p}, N(1-N/2^n)/4)$ under the assumption that $(\epsilon + \theta_0)^2$ is negligible. Again, we note that the approximation holds in the mentioned range of $\mathfrak{p}$ and it is not valid for values of $\mathfrak{p}$ close to 0 or 1.

$$\begin{aligned}
\Pr[X_{\kappa^*} \leq x] &\approx \int_{p-\theta_0}^{p+\theta_0} \left(\int_{-\infty}^{x} \mathfrak{f}(x; N\mathfrak{p}, N(1 - N/2^n)/4)\ dx \right) \mathfrak{f}(\mathfrak{p}; p, s_0^2)\ d\mathfrak{p} \\
&\leq \int_{-\infty}^{\infty} \left(\int_{-\infty}^{x} \mathfrak{f}(x; N\mathfrak{p}, N(1 - N/2^n)/4)\ dx \right) \mathfrak{f}(\mathfrak{p}; p, s_0^2)\ d\mathfrak{p} \\
&= \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} \mathfrak{f}(x; N\mathfrak{p}, N(1 - N/2^n)/4) \mathfrak{f}(\mathfrak{p}; p, s_0^2)\ d\mathfrak{p} \right) dx \\
&= \int_{-\infty}^{x} \mathfrak{f}(x; Np, s_0^2 N^2 + N(1 - N/2^n)/4) dx.
\end{aligned}$$

The last equality follows from Proposition 1 in Section A.2. So, X_{κ^*} approximately follows $\mathcal{N}(Np, s_0^2 N^2 + N(1 - N/2^n)/4)$ and since $W_{\kappa^*} = X_{\kappa^*}/N - 1/2$ we have that the distribution of W_{κ^*} is approximately given as follows:

$$W_{\kappa^*} \sim \mathcal{N}\left(\epsilon, s_0^2 + \frac{1 - N/2^n}{4N}\right). \tag{33}$$

For W_{κ} with $\kappa \neq \kappa^*$, we need to consider the general wrong key randomisation hypothesis where p_{κ,κ^*} is modelled as a random variable following $\mathcal{N}(1/2, s_1^2)$. In this case, it is required to use (22) and (23) instead

of (20) and (21) respectively. In particular, as in the case of sampling with replacement, we note that for $\mathfrak{p} \in [1/2 - \vartheta_1, 1/2 + \vartheta_1]$, it is required to approximate $\mathcal{N}(N\mathfrak{p}, N\mathfrak{p}(1-\mathfrak{p}))$ by $\mathcal{N}(N/2, N/4)$, i.e., $\mathfrak{p}(1-\mathfrak{p}) \approx 1/4$. The validity of this follows from (23) and the approximation is not valid for values of $\mathfrak{p}$ near to 0 or 1. With these approximations, the resulting analysis shows the following approximate distribution:

$$W_\kappa \;\sim\; \mathcal{N}\left(0, s_1^2 + \frac{1 - N/2^n}{4N}\right), \quad \kappa \neq \kappa^*. \tag{34}$$

Remark: In [1], for the adjusted wrong key randomisation hypothesis, i.e., with $s_1^2 = 2^{-n-2}$, the distribution of W_κ for $\kappa \neq \kappa^*$ was stated to be $\mathcal{N}\left(0, \frac{1}{4N}\right)$. We note the following issues.

1. The supporting argument in [1] was given to be the fact that if two random variables X and Y are such that $X \sim \mathcal{N}(\mathfrak{a}Y, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu, \sigma_2^2)$, then $X \sim \mathcal{N}(\mathfrak{a}\mu, \sigma_1^2 + \mathfrak{a}^2\sigma_2^2)$ (see Proposition 2 in the appendix for a proof). This argument, however, is not complete. The distribution function of X_κ for $\kappa \neq \kappa^*$ is

$$\Pr[X_{\kappa^*} \leq x] \;=\; \sum_{\mathfrak{k} \leq x} \Pr[X_\kappa = \mathfrak{k}] = \sum_{\mathfrak{k} \leq x} \int_{-\infty}^{\infty} \frac{\binom{2^{n-1}}{\mathfrak{k}}\binom{2^n - 2^{n-1}}{N-\mathfrak{k}}}{\binom{2^n}{N}} \mathfrak{f}(\mathfrak{p}; 1/2, s_1^2) d\mathfrak{p}. \tag{35}$$

   After interchanging the order of the sum and the integration, one can apply the normal approximation of the hypergeometric distribution. It is not justified to directly start with the normal approximation of the hypergeometric distribution as has been done in [1].

2. The issue is more subtle than simply a question of interchanging the order of the sum and the integral. After applying the normal approximation of the hypergeometric distribution one ends up with $\mathcal{N}(N/2, N(1 - N/2^n)\mathfrak{p}(1-\mathfrak{p}))$ which is then approximated as $\mathcal{N}(N/2, N(1 - N/2^n)/4)$. This requires assuming that $(\mathfrak{p} - 1/2)^2$ is negligible. Clearly, this assumption is not valid for values of $\mathfrak{p}$ close to 0 or 1. On the other hand, the approximation is justified for $\mathfrak{p} \in [1/2 - \vartheta_1, 1/2 + \vartheta_1]$ under the assumption that $s_1^2 2^{n/2} = 2^{-2-n/2}$ is negligible (see (23)). Also, the probability that $\mathfrak{p}$ takes values outside $[1/2 - \vartheta_1, 1/2 + \vartheta_1]$ is exponentially low as shown in (22). So, it is required to argue that the integral in (35) is from $1/2 - \vartheta_1$ to $1/2 + \vartheta_1$ and the contribution of the integral outside this range is negligible. This can be done in a manner which is similar to that done in Steps (25) to (26). In [1], the assumption that $(\mathfrak{p} - 1/2)^2$ is negligible has been made for all values of $\mathfrak{p}$ which is not justified.

5.3 Success Probability under General Key Randomisation Hypotheses

The distributions of W_{κ^*} and W_κ for $\kappa \neq \kappa^*$ are respectively given by (30) and (31) for the case of sampling with replacement and are given by (33) and (34) for the case of sampling without replacement. These expressions can be compactly expressed in the following form:

$$W_{\kappa^*} \sim \mathcal{N}(\epsilon, s_0^2 + \sigma^2); \qquad W_\kappa \sim \mathcal{N}(0, s_1^2 + \sigma^2), \quad \text{for } \kappa \neq \kappa^*; \tag{36}$$

where

$$\sigma^2 \;=\; \begin{cases} \frac{1}{4N} & \text{for sampling with replacement;} \\ \frac{1 - N/2^n}{4N} & \text{for sampling without replacement.} \end{cases} \tag{37}$$

Substituting $\sigma_0^2 = s_0^2 + \sigma^2$ and $\sigma_1^2 = s_1^2 + \sigma^2$ in Theorem 1, we obtain the following result.

Theorem 2. *Let $\kappa^* \in \{0,1\}^m$. For $\kappa \in \{0,1\}^m$, let $T_\kappa = |W_\kappa|$ be 2^m random variables, where $W_{\kappa^*} \sim \mathcal{N}(\mu_0, s_0^2 + \sigma^2)$, $\mu_0 \neq 0$ and $W_\kappa \sim \mathcal{N}(0, s_1^2 + \sigma^2)$ for $\kappa \neq \kappa^*$. Suppose the hypothesis test given in (12) is applied to T_κ for all $\kappa \in \{0,1\}^m$. Let $P_S = 1 - \Pr[\text{Type-1 error}]$. Then*

$$P_S = \Phi\left(\frac{|\epsilon| - \sqrt{s_1^2 + \sigma^2}\gamma}{\sqrt{s_0^2 + \sigma^2}}\right) + \Phi\left(\frac{-|\epsilon| - \sqrt{s_1^2 + \sigma^2}\gamma}{\sqrt{s_0^2 + \sigma^2}}\right) \tag{38}$$

where $\gamma = \Phi^{-1}\left(1 - \frac{2^{m-a-1}}{2^m - 1}\right)$ and the expected number of times that Type-2 errors occurs is 2^{m-a}.

Let $P_S^{(\mathrm{wr})}$ denote the success probability when sampling with replacement is used and let $P_S^{(\mathrm{wor})}$ denote the success probability when sampling without replacement is used. Using the corresponding expressions for σ from (37) in (38) we obtain the following expressions for $P_S^{(\mathrm{wr})}$ and $P_S^{(\mathrm{wor})}$.

$$P_S^{(\mathrm{wr})} = \Phi\left(\frac{2\sqrt{N}|\epsilon| - \sqrt{1 + 4Ns_1^2}\gamma}{\sqrt{1 + 4Ns_0^2}}\right) + \Phi\left(\frac{-2\sqrt{N}|\epsilon| - \sqrt{1 + 4Ns_1^2}\gamma}{\sqrt{1 + 4Ns_0^2}}\right); \tag{39}$$

$$P_S^{(\mathrm{wor})} = \Phi\left(\frac{2\sqrt{N}|\epsilon| - \sqrt{4Ns_1^2 + (1 - N/2^n)}\gamma}{\sqrt{4Ns_0^2 + (1 - N/2^n)}}\right) + \Phi\left(\frac{-2\sqrt{N}|\epsilon| - \sqrt{4Ns_1^2 + (1 - N/2^n)}\gamma}{\sqrt{4Ns_0^2 + (1 - N/2^n)}}\right). \tag{40}$$

Remarks:

1. If $N \ll 2^n$, then $P_S^{(\mathrm{wor})} \approx P_S^{(\mathrm{wr})}$. So, the expression for $P_S^{(\mathrm{wor})}$ given by (40) becomes useful only when the fraction $N/2^n$ is non-negligible.
2. In the case of sampling with replacement, due to the birthday paradox, having N to be greater than $2^{n/2}$ is not really useful, since repetitions will begin to occur.

5.4 Success Probability Under Standard Key Randomisation Hypotheses

Let $P_S^{(\mathrm{wr,std})}$ and $P_S^{(\mathrm{wor,std})}$ be the success probabilities for standard key randomisation hypotheses corresponding to the situations where plaintexts are chosen with and without replacement respectively. As discussed in Section 4, the standard key randomisation hypotheses is obtained from the general key randomisation hypothesis by letting $s_0 \downarrow 0$ and $s_1 \downarrow 0$. Using these conditions in (39) and (40) lead to the following expressions for the success probabilities in the two cases of sampling with and without replacement.

$$P_S^{(\mathrm{wr,std})} = \Phi\left(2\sqrt{N}|\epsilon| - \gamma\right) + \Phi\left(-2\sqrt{N}|\epsilon| - \gamma\right). \tag{41}$$

$$P_S^{(\mathrm{wor,std})} = \Phi\left(\frac{2\sqrt{N}}{\sqrt{1 - N/2^n}}|\epsilon| - \gamma\right) + \Phi\left(-\frac{2\sqrt{N}}{\sqrt{1 - N/2^n}}|\epsilon| - \gamma\right). \tag{42}$$

Success probability in [26]: Selçuk [26] had obtained an expression for the success probability under the standard key randomisation hypotheses and under the assumption that $P_1, \ldots, P_N$ are chosen uniformly with replacements. The expression for $P_S^{(\mathrm{wr,std})}$ given by (41) was not obtained in [26]. This is due to the following reasons.

1. For analysing the success probability, Selçuk [26] employed the order statistics based approach. As discussed in Section 3.1, in this approach the T's are written as $T_0, \ldots, T_{2^m-1}$ and it is assumed that T_0 corresponds to the right key. With this set-up, an attack with a-bit advantage is successful, if $T_0 > T_{(2^m q)}$ where $q = 1 - 2^{-a}$. Selçuk [26] insteads considers success to be the event $W_0 > T_{(2^m q)}$ and the condition $W_0/\mu_0 > 0$. Since

the T's can take only non-negative values, it follows that $T_{(2^m q)} \geq 0$ and so the event $W_0 > T_{(2^m q)}$ implies $W_0 > 0$ and so $\mu_0 > 0$. Conversely, if $\mu_0 < 0$, then for the condition $W_0/\mu_0 > 0$ to hold we must have $W_0 < 0$ in which case the event $W_0 > T_{(2^m q)}$ is an impossible event. So, the condition $W_0 > T_{(2^m q)}$ subsumes the condition $W_0/\mu_0 > 0$ for $\mu_0 > 0$ and has probability 0 for $\mu_0 < 0$. No justification is provided in [26] for considering success to be $W_0 > T_{(2^m q)}$ instead of $T_0 > T_{(2^m q)}$. From (10) we see that the event $W_0 > T_{(2^m q)}$ is a sub-event of $T_0 > T_{(2^m q)}$ which is the event that the attack is successful.

2. It is assumed that $\sigma_0 \gg \sigma_q$. This is justified in [26] by providing numerical values for a in the range $8 \leq a \leq 48$ and it is mentioned that the assumption especially holds for success probability 0.8 or more.

Under the above two assumptions, the expression for success probability obtained in [26] is the following.

$$P_S \approx \Phi\left(2\sqrt{N}|\epsilon| - \Phi^{-1}(1 - 2^{-a-1})\right). \tag{43}$$

Assume that m is large so that $2^m - 1 \approx 2^m$ and so $\gamma \approx \Phi^{-1}(1 - 2^{-a-1})$. Then the right hand side of (43) becomes equal to the first term of (41). This shows that the expression for the success probability obtained in [26] is incomplete.

To the best of our knowledge, no prior work has analysed the success probability of linear cryptanalysis with the standard key randomisation hypotheses and under the condition where $P_1, \ldots, P_N$ are chosen uniformly without replacement. So, the expression for $P_S^{(\text{wor,std})}$ given by (42) is the first such result.

5.5 Success Probability Under Adjusted Wrong Key Randomisation Hypothesis

Let $P_S^{(\text{wr,adj})}$ and $P_S^{(\text{wor,adj})}$ be the success probabilities for adjusted wrong key randomisation hypothesis (and standard right key randomisation hypothesis) corresponding to the situations where plaintexts are chosen with and without replacement respectively.

Setting $s_1^2 = 2^{-n-2}$ converts the general wrong key randomisation hypothesis to the adjusted wrong key randomisation hypothesis. Also, we let $s_0 \downarrow 0$, so that the general right key randomisation hypothesis simplifies to the standard right key randomisation hypothesis. Using the conditions for s_0 and s_1 in (39) and (40) provides the following expressions for the success probabilities in the two cases of sampling with and without replacement.

$$P_S^{(\text{wr,adj})} = \Phi\left(2\sqrt{N}|\epsilon| - \sqrt{1 + N/2^n}\gamma\right) + \Phi\left(-2\sqrt{N}|\epsilon| - \sqrt{1 + N/2^n}\gamma\right). \tag{44}$$

$$P_S^{(\text{wor,adj})} = \Phi\left(\frac{2\sqrt{N}|\epsilon| - \gamma}{\sqrt{1 - N/2^n}}\right) + \Phi\left(\frac{-2\sqrt{N}|\epsilon| - \gamma}{\sqrt{1 - N/2^n}}\right). \tag{45}$$

Expressions for the success probability with the adjusted wrong key randomisation hypothesis and the standard right key randomisation hypothesis were obtained in [8] and [1]. Both the works followed the order statistics approach as used by Selçuk. The work [8] considered the setting of uniform random choice of $P_1, \ldots, P_N$ with replacement whereas [1] considered the setting of uniform random choice of $P_1, \ldots, P_N$ without replacement. Under the approximation $2^m \approx 2^m - 1$, the expressions obtained in [8] and [1] are equal to the first terms of (44) and (45) respectively. The reason why the complete expressions were not obtained in [8, 1] is similar to the reason why Selçuk was not able to obtain the complete expression for $P_S^{(\text{wr,std})}$.

The expressions for both $P_S^{(\text{wr,adj})}$ and $P_S^{(\text{wor,adj})}$ can be seen to be functions of $|\epsilon|$, N and γ. Since γ itself is a function of the advantage a and the size of the target sub-key m, it follows that both $P_S^{(\text{wr,adj})}$ and $P_S^{(\text{wor,adj})}$ are functions of $|\epsilon|$, N, a and m. None of these quantities are random variables, so neither are $P_S^{(\text{wr,adj})}$ and $P_S^{(\text{wor,adj})}$ random variables. Consequently, it is not meaningful to talk about the average value of P_S or about the probability that P_S is monotonous as has been done in [1].

6 Understanding Non-Monotonic Behaviour

As in Section 3, let $W_{\kappa^*} \sim \mathcal{N}(\mu_0, \sigma_0^2)$ and $W_\kappa \sim \mathcal{N}(0, \sigma_1^2)$ for $\kappa \neq \kappa^*$. Assume $\mu_0 > 0$. From (17), we obtain the expression for α, the probability of Type-1 error, to be $\alpha = 1 - P_S = \Phi((t-\mu_0)/\sigma_0) + \Phi((-t-\mu_0)/\sigma_0)$. The first term arises from the upper bound on W_{κ^*} given in (16), i.e.,

$$\Pr[W_{\kappa^*} \leq t] \quad = \quad \Phi((t-\mu_0)/\sigma_0). \tag{46}$$

From (19), $t = \sigma_1\gamma$, where γ is a constant. Let $\pi = \Pr[W_{\kappa^*} \leq t]$ and then $1-\pi$ is the corresponding contribution to P_S. Note that π is the area under the curve of the density function of W_{κ^*} from $-\infty$ to t.

Consider the setting of general key randomisation hypothesis where $P_1, \ldots, P_N$ are chosen with replacement. From (30) and (31), we have $\mu_0 = \epsilon$, $\sigma_0^2 = s_0^2 + \frac{1}{4N}$ and $\sigma_1^2 = s_1^2 + \frac{1}{4N}$. Since s_0 and s_1 are constants, σ_0 and σ_1 are both inversely proportional to N. So, as N increases the normal curve for W_{κ^*} becomes more concentrated around the mean ϵ. This is shown in Figures 1, 2 and 3. Also, since γ is a constant, $t = \sigma_1\gamma$ is also inversely proportional to N. So, π is a function of N. One may expect π to be a monotonic decreasing function of N (and $1-\pi$ to be a monotonic increasing function of N), but, this does not necessarily hold as we explain below.

Let $N_1 < N_2$. The corresponding density functions for W_{κ^*} are $\mathfrak{f}(x; \epsilon, s_0^2 + 1/(4N_1))$ and $\mathfrak{f}(x; \epsilon, s_0^2 + 1/(4N_2))$. So, there is an x_0 such that the following hold:

- $\mathfrak{f}(x; \epsilon, s_0^2 + 1/(4N_1)) \geq \mathfrak{f}(x; \epsilon, s_0^2 + 1/(4N_2))$, for $x \geq x_0$;
- $\mathfrak{f}(x; \epsilon, s_0^2 + 1/(4N_1)) < \mathfrak{f}(x; \epsilon, s_0^2 + 1/(4N_2))$, for $x < x_0$.

The point x_0 is shown in Figures 1, 2 and 3.

Let $t_1 = (s_1^2 + 1/(4N_1))\gamma$ and $t_2 = (s_1^2 + 1/(4N_2))\gamma$ and so $t_2 < t_1$. Let π_1 and π_2 be the values of π corresponding to N_1 and N_2. There are two possibilities.

$t_2 \leq x_0$: In this case, we have either $t_2 < t_1 \leq x_0$ or $t_2 \leq x_0 < t_1$. From Figures 1 and 2, in both cases, it can be noted that the area under the curve corresponding to N_1 is more than the area under the curve corresponding to N_2. So, $\pi_1 > \pi_2$. In other words, increasing N leads to π going down and correspondingly $1-\pi$ going up. As a result, in this case, the first term in the expression for success probability given by (17) increases with N.

$t_2 > x_0$: In this case, we have $x_0 < t_2 < t_1$. From Figure 3, it is no longer clear that the area under the curve corresponding to N_1 is more than the area under the curve corresponding to N_2. So, it cannot be definitely said that π_1 is more than π_2 and so the $1-\pi$ does not necessarily go up. As a result, it can no longer be said that the first term in the expression for success probability given by (17) increases with N.

Note that the above explanation is purely statistical in nature. It is entirely based upon the expressions for the variances of the two normal distributions.

In the above discussion, we have tried to explain the possible non-monotonic behaviour of the probability of the event $\Pr[W_{\kappa^*} \leq t]$ for the case of sampling with replacement. Considering this specific case makes it easy to see the dependence of the variances on N in determining possible non-monotonicity. The explanation extends to the complete expression for the success probability as well as to the case of sampling without replacement.

Explanations for non-monotonic behaviour have been provided in [8, 1]. In [8], non-monotonicity has essentially been attributed to the strategy of sampling with replacement leading to duplicates. The later work [1], observed non-monotonicity even for the strategy of sampling without replacement and so the explanation based on the occurrence of duplicates could not be applied. Instead, [1] provides an explanation for non-monotonicity for both sampling with and without replacement based on the ranking strategy used in the order statistics based approach. As we have seen, expressions for success probability can be obtained without using the order statistics based approach. So, an explanation of non-monotonicity based on order statistics based approach is not adequate. Instead, as we have tried to explain above, the phenomenon is better understood by considering that the variances of the two normal distributions in question are monotone decreasing with N.

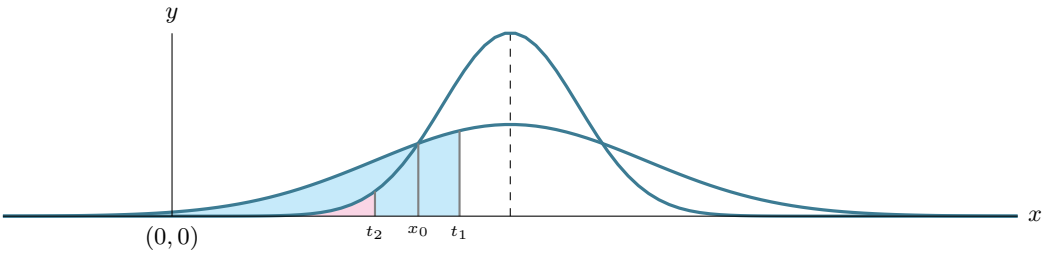

Figure 1: Case $t_2 \leq x_0 < t_1$.

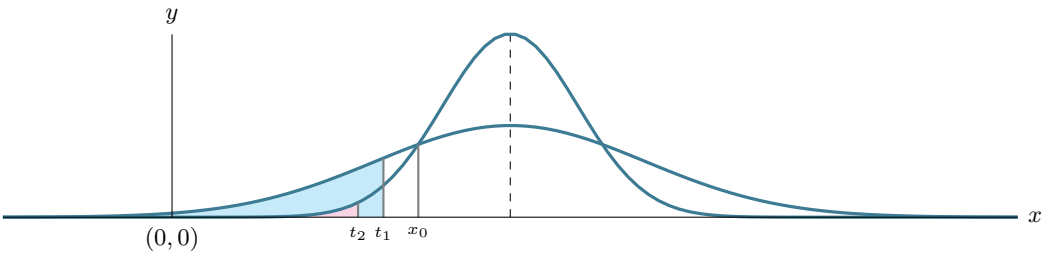

Figure 2: Case $t_2 < t_1 < x_0$.

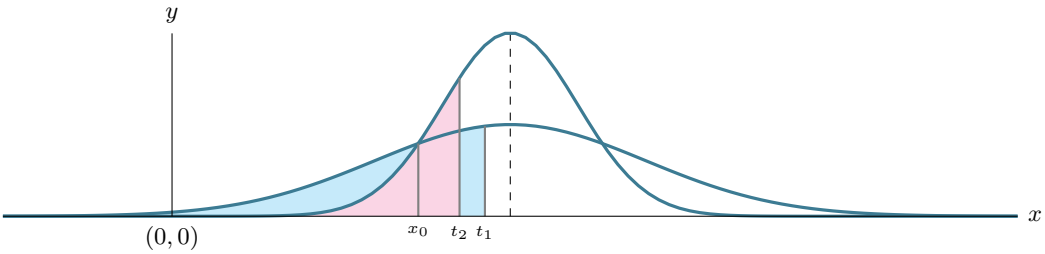

Figure 3: Case $x_0 < t_2 < t_1$.

6.1 Dependence of P_S on N

Consider the general expression for the success probability P_S as given by (38). The subsequent expressions for success probability with/without replacement and under standard/adjusted key randomisation hypotheses are all obtained as special cases of (38). In (38), the quantities s_0, s_1 and γ are constants which are independent of N and only σ depends on N as shown in (37). Further, from (37), it is clear that σ is a decreasing function of N for both the cases of with and without replacements.

We analyse the behaviour of P_S as a function of N and identify the situations where P_S is a monotonic increasing function of N.

Theorem 3. *Consider P_S to be given by (38) where s_0, s_1 and γ are positive and independent of N while $\sigma > 0$ is a monotone decreasing function of N.*

1. *Suppose $s_0 \geq s_1$. Then P_S is an increasing function of N for all $N > 0$.*
2. *Suppose $s_0 < s_1$ and $\left(s_1^2 - s_0^2\right)\gamma \geq |\epsilon|\sqrt{\sigma^2 + s_1^2}$. Then P_S is a decreasing function of N.*
3. *Suppose $s_0 < s_1$, $\left(s_1^2 - s_0^2\right)\gamma < |\epsilon|\sqrt{\sigma^2 + s_1^2}$ and $\delta = \frac{\left(s_1^2 - s_0^2\right)\gamma}{|\epsilon|\sqrt{\sigma^2+s_1^2}}$ is such that δ^3 and higher powers of δ can be ignored. Then P_S is an increasing function of N if and only if $\sigma^2\left(\left(s_1^2 - s_0^2\right) - \epsilon^2\right) < \epsilon^2 s_1^2 - \left(s_1^2 - s_0^2\right)s_0^2$.*

Proof. We proceed by taking derivatives with respect to N. Since σ is a decreasing function of N, $\frac{d\sigma}{dN} < 0$.

$$\begin{aligned}
\frac{dP_S}{dN} &= \phi\left(\frac{|\epsilon| - \sqrt{\sigma^2 + s_1^2}\gamma}{\sqrt{\sigma^2 + s_0^2}}\right)\left(-\frac{\gamma\sigma\frac{d\sigma}{dN}}{\sqrt{\sigma^2 + s_1^2}} \cdot \frac{1}{\sqrt{\sigma^2 + s_0^2}} + \left(|\epsilon| - \sqrt{\sigma^2 + s_1^2}\gamma\right) \cdot \left(-\frac{\sigma\frac{d\sigma}{dN}}{\sqrt{\left(\sigma^2 + s_0^2\right)^3}}\right)\right) - \\
&\quad \phi\left(-\frac{|\epsilon| + \sqrt{\sigma^2 + s_1^2}\gamma}{\sqrt{\sigma^2 + s_0^2}}\right)\left(\frac{\gamma\sigma\frac{d\sigma}{dN}}{\sqrt{\sigma^2 + s_1^2}} \cdot \frac{1}{\sqrt{\sigma^2 + s_0^2}} + \left(|\epsilon| + \sqrt{\sigma^2 + s_1^2}\gamma\right) \cdot \left(-\frac{\sigma\frac{d\sigma}{dN}}{\sqrt{\left(\sigma^2 + s_0^2\right)^3}}\right)\right) \\
&= \frac{\sigma f_1(\sigma)\frac{d\sigma}{dN}}{\sqrt{\sigma^2 + s_1^2}\sqrt{\left(\sigma^2 + s_0^2\right)^3}}, \text{ where} \\
f_1(\sigma) &= \phi\left(\frac{|\epsilon| - \sqrt{\sigma^2 + s_1^2}\gamma}{\sqrt{\sigma^2 + s_0^2}}\right)\left(-\gamma\left(\sigma^2 + s_0^2\right) - \left(|\epsilon|\sqrt{\sigma^2 + s_1^2} - \left(\sigma^2 + s_1^2\right)\gamma\right)\right) \\
&\quad -\phi\left(-\frac{|\epsilon| + \sqrt{\sigma^2 + s_1^2}\gamma}{\sqrt{\sigma^2 + s_0^2}}\right)\left(\gamma\left(\sigma^2 + s_0^2\right) - \left(|\epsilon|\sqrt{\sigma^2 + s_1^2} + \left(\sigma^2 + s_1^2\right)\gamma\right)\right).
\end{aligned}$$

Using the definition of the standard normal density function, we have

$$\begin{aligned}
f_1(\sigma) &= \frac{e^{\frac{|\epsilon| + \sqrt{\sigma^2 + s_1^2}\gamma}{2(\sigma^2 + s_0^2)}}}{\sqrt{2\pi}}\left[-\left(\sigma^2 + s_0^2\right)\gamma\left(e^{\frac{2|\epsilon|\gamma\sqrt{\sigma^2+s_1^2}}{\sigma^2+s_0^2}} + 1\right) - |\epsilon|\sqrt{\sigma^2 + s_1^2}\left(e^{\frac{2|\epsilon|\gamma\sqrt{\sigma^2+s_1^2}}{\sigma^2+s_0^2}} - 1\right) + \right. \\
&\quad \left.\left(\sigma^2 + s_1^2\right)\gamma\left(e^{\frac{2|\epsilon|\gamma\sqrt{\sigma^2+s_1^2}}{\sigma^2+s_0^2}} + 1\right)\right] \\
&= \frac{-1}{\sqrt{2\pi}}\exp\left(\frac{|\epsilon| + \sqrt{\sigma^2 + s_1^2}\gamma}{2(\sigma^2 + s_0^2)}\right) f_2(\sigma), \text{ where} \\
f_2(\sigma) &= \left(s_0^2 - s_1^2\right)\gamma\left(e^{\frac{2|\epsilon|\gamma\sqrt{\sigma^2+s_1^2}}{\sigma^2+s_0^2}} + 1\right) + |\epsilon|\sqrt{\sigma^2 + s_1^2}\left(e^{\frac{2|\epsilon|\gamma\sqrt{\sigma^2+s_1^2}}{\sigma^2+s_0^2}} - 1\right).
\end{aligned}$$

Since $d\sigma/dN < 0$ we have that $dP_S/dN > 0$ if and only if $f_1(\sigma) < 0$ if and only if $f_2(\sigma) > 0$. If $s_0 \geq s_1$, then $f_2(\sigma) > 0$ and so in this case we have $dP_S/dN > 0$ which implies that P_S is an increasing function of N. This proves the first point.

Now consider the case $s_0 < s_1$. We write

$$\begin{aligned} f_2(\sigma) &= -\left(s_1^2 - s_0^2\right)\gamma \left(e^{\frac{2|\epsilon|\gamma\sqrt{\sigma^2+s_1^2}}{\sigma^2+s_0^2}} + 1 \right) + |\epsilon|\sqrt{\sigma^2+s_1^2} \left(e^{\frac{2|\epsilon|\gamma\sqrt{\sigma^2+s_1^2}}{\sigma^2+s_0^2}} - 1 \right). \\ &= e^{\frac{2|\epsilon|\gamma\sqrt{\sigma^2+s_1^2}}{\sigma^2+s_0^2}} \left(|\epsilon|\sqrt{\sigma^2+s_1^2} - \left(s_1^2 - s_0^2\right)\gamma \right) - |\epsilon|\sqrt{\sigma^2+s_1^2} - \left(s_1^2 - s_0^2\right)\gamma. \end{aligned}$$

If $\left(s_1^2 - s_0^2\right)\gamma \geq |\epsilon|\sqrt{\sigma^2+s_1^2}$, then $f_2(\sigma) < 0$ and so $dP_S/dN < 0$ which implies that P_S is a decreasing function of N. This proves the second point.

So, suppose that $s_0 < s_1$ and $\left(s_1^2 - s_0^2\right)\gamma < |\epsilon|\sqrt{\sigma^2+s_1^2}$ both hold. By the condition of this case, we have $0 < \delta < 1$. Also, we have the assumption that δ is small enough such that δ^3 and higher powers of δ can be ignored. Then $f_2(\sigma) > 0$ if and only if

$$\frac{2|\epsilon|\gamma\sqrt{\sigma^2+s_1^2}}{\sigma^2+s_0^2} > \ln\left(\frac{1+\delta}{1-\delta}\right) \approx 2\delta = 2\frac{\left(s_1^2 - s_0^2\right)\gamma}{|\epsilon|\sqrt{\sigma^2+s_1^2}}.$$

Cancelling 2γ on both sides and rearranging the terms shows the third point. □

Fisher information: Suppose a random variable Y follows a distribution whose density is given by a function $\mathfrak{g}(y;\theta_1,\theta_2,\ldots)$, where $\theta_1,\theta_2,\ldots$ are the finitely many parameters specifying the density function. A relevant question is how much information does the random variable Y carry about one particular parameter θ_i. Fisher information is a well known measure in statistics for quantifying this information. The Fisher information about a parameter $\theta \in \{\theta_1,\theta_2,\ldots\}$ carried in the random variable Y is defined to be

$$\mathcal{I}_Y(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\ln\mathfrak{g}(Y;\theta_1,\theta_2,\ldots)\right)^2\right]. \tag{47}$$

If $Y \sim \mathcal{N}(\mu,\sigma^2)$, then $\mathcal{I}_Y(\mu) = \sigma^{-2}$. In other words, the information contained in the random variable Y is inversely proportional to σ^2. So, as the variance increases, the information about the mean contained in the random variable Y decreases.

We view the first point of Theorem 3 in the context of Fisher information. Recall that p_{κ^*} is a random variable following $\mathcal{N}(p,s_0^2)$ and p_{κ,κ^*} is a random variable following $\mathcal{N}(1/2,s_1^2)$. So, $\mathcal{I}_{p_{\kappa^*}}(p) = s_0^{-2}$ and $\mathcal{I}_{p_{\kappa,\kappa^*}}(1/2) = s_1^{-2}$. From the first point of Theorem 3 we have that if $s_0 > s_1$, then P_S is an increasing function of N for all $N > 0$. Put in terms of Fisher information, this is equivalent to saying that if $\mathcal{I}_{p_{\kappa,\kappa^*}}(1/2) \geq \mathcal{I}_{p_{\kappa^*}}(p)$, then P_S is an increasing function of N. More explicitly, if the information about the mean contained in p_{κ^*} is not more than the information about the mean contained in p_{κ,κ^*}, then increasing N increases the success probability. Viewed differently, if the variability of p_{κ^*} is at least as much as the variability of p_{κ,κ^*}, then the chances of the attack being successful increases as the number of observations increases.

Applying Theorem 3 to the case of standard key randomisation hypothesis, we have $s_0 \downarrow 0$ and $s_1 \downarrow 0$. So, by the first point of Theorem 3, it follows that both $P_S^{(\mathrm{wr,std})}$ and $P_S^{(\mathrm{wor,std})}$ are increasing functions of N for all $N > 0$.

6.2 Adjusted Wrong Key Randomisation Hypothesis

In this case $s_1^2 = 2^{-n-2}$. Also, assuming the standard right key randomisation hypothesis (as in [8, 1]), $s_0 \downarrow 0$. So, Points 2 and 3 of Theorem 3 apply. This case is divided into two subcases.

Sampling with replacement: In this case, $\sigma^2 = 1/(4N)$. Let $N_0^{(\mathrm{wr})} = (s_1^2 - \epsilon^2)/(4\epsilon^2 s_1^2)$ and note that $N_0^{(\mathrm{wr})} > 0$ if and only if $s_1 < |\epsilon|$.

1. Suppose $s_1\gamma > |\epsilon|$ so that $s_1^4\gamma^2 - \epsilon^2 s_1^2 > 0$.

   (a) By Point 2 of Theorem 3, for $N \geq \epsilon^2/(4(s_1^4\gamma^2 - \epsilon^2 s_1^2))$, $P_S^{(\mathrm{wr,adj})}$ is a decreasing function of N.

   (b) By Point 3 of Theorem 3, for $(s_1^2 - \epsilon^2)/(4\epsilon^2 s_1^2) < N < \epsilon^2/(4(s_1^4\gamma^2 - \epsilon^2 s_1^2))$, $P_S^{(\mathrm{wr,adj})}$ is an increasing function of N and for $N < (s_1^2 - \epsilon^2)/(4\epsilon^2 s_1^2)$, $P_S^{(\mathrm{wr,adj})}$ is a decreasing function of N.

   Recall that N is the number of plaintext-ciphertext pairs and hence is positive and at most 2^n. Let $N_1^{(\mathrm{wr})} = \epsilon^2/(4(s_1^4\gamma^2 - \epsilon^2 s_1^2))$. We have $s_1^2 = 2^{-n-2}$ and so, $N_1^{(\mathrm{wr})} < 2^n$ if and only if $|\epsilon| < (s_1\gamma)/\sqrt{2}$. For sampling with replacement, it is more meaningful to consider $2^{n/2}$ to be the upper bound for N, since beyond a sample size of $2^{n/2}$ there will be too many repetitions in the sample. We have $N_1^{(\mathrm{wr})} < 2^{n/2}$ if and only if $|\epsilon| < s_1\gamma/\sqrt{1 + (2s_1)^{-1}}$. This means that if $|\epsilon| < (s_1\gamma)/\sqrt{2}$, $P_S^{(\mathrm{wr,adj})}$ is a decreasing function of N for $N_1^{(\mathrm{wr})} \leq N \leq 2^n$. So, for $|\epsilon| < s_1\gamma/\sqrt{1 + (2s_1)^{-1}}$, $P_S^{(\mathrm{wr,adj})}$ is a decreasing function of N for $N_0^{(\mathrm{wr})} \leq N \leq 2^{n/2}$.

2. Suppose $s_1\gamma < |\epsilon|$ so that $s_1^4\gamma^2 - \epsilon^2 s_1^2 < 0$.

   (a) By Point 2 of Theorem 3, for $N \leq -\epsilon^2/(4(\epsilon^2 s_1^2 - s_1^4\gamma^2))$, $P_S^{(\mathrm{wr,adj})}$ is a decreasing function of N.

   (b) By Point 3 of Theorem 3, for $N > (s_1^2 - \epsilon^2)/(4\epsilon^2 s_1^2)$, $P_S^{(\mathrm{wr,adj})}$ is an increasing function of N and for $-\epsilon^2/(4(\epsilon^2 s_1^2 - s_1^4\gamma^2)) < N < (s_1^2 - \epsilon^2)/(4\epsilon^2 s_1^2)$, $P_S^{(\mathrm{wr,adj})}$ is a decreasing function of N.

   We have the following.

The above is summarised as follows:

Case $|\epsilon| < \min(s_1, (s_1\gamma)/\sqrt{2}) < s_1\gamma$**:**

- $P_S^{(\mathrm{wr,adj})}$ is a decreasing function of N in the range $0 < N < N_0^{(\mathrm{wr})}$;
- $P_S^{(\mathrm{wr,adj})}$ is an increasing function of N in the range $N_0^{(\mathrm{wr})} < N < N_1^{(\mathrm{wr})}$; and
- $P_S^{(\mathrm{wr,adj})}$ is a decreasing function of N in the range $N_1^{(\mathrm{wr})} < N < 2^{n/2}$.
- $P_S^{(\mathrm{wr,adj})}$ attains a minima at $N_0^{(\mathrm{wr})}$ and a maximum at $N_1^{(\mathrm{wr})}$.

Case $s_1 < |\epsilon| < (s_1\gamma)/\sqrt{2} < s_1\gamma$**:**

- $P_S^{(\mathrm{wr,adj})}$ is an increasing function of N in the range $0 < N < N_1$; and
- $P_S^{(\mathrm{wr,adj})}$ is a decreasing function of N in the range $N_1^{(\mathrm{wr})} < N < 2^{n/2}$.
- $P_S^{(\mathrm{wr,adj})}$ attains a maximum at $N_1^{(\mathrm{wr})}$.

Case $s_1\gamma < |\epsilon| < s_1$**:**

- $P_S^{(\mathrm{wr,adj})}$ is a decreasing function of N for $0 < N < N_0^{(\mathrm{wr})}$; and
- $P_S^{(\mathrm{wr,adj})}$ is an increasing function of N for $N > N_0^{(\mathrm{wr})}$.
- $P_S^{(\mathrm{wr,adj})}$ attains a minima at $N_0^{(\mathrm{wr})}$.

Case $\max(s_1, s_1\gamma) < |\epsilon|$**:** $P_S^{(\mathrm{wr,adj})}$ is an increasing function of N for $N > 0$.

Sampling without replacement: In this case, $\sigma^2 = 1/(4N) - 1/2^{-n-2} = 1/(4N) - s_1^2$ and so $\sigma^2 + s_1^2 = 1/(4N)$. Let $N_0^{(\text{wor})} = (s_1^2 - \epsilon^2)/(4s_1^4)$ and so $N_0^{(\text{wr})} > 0$ if and only if $|\epsilon| < s_1$.

1. By Point 2 of Theorem 3, for $N \geq \epsilon^2/(4s_1^4\gamma^2)$, $P_S^{(\text{wor,adj})}$ is a decreasing function of N.
2. By Point 3 of Theorem 3, for $(s_1^2 - \epsilon^2)/(4s_1^4) < N < \epsilon^2/(4s_1^4\gamma^2)$, $P_S^{(\text{wor,adj})}$ is an increasing function of N.
3. By Point 3 of Theorem 3, for $N < (s_1^2 - \epsilon^2)/(4s_1^4)$, $P_S^{(\text{wor,adj})}$ is a decreasing function of N.

Let $N_1^{(\text{wor})} = \epsilon^2/(4s_1^4\gamma^2)$ and so $N_1^{(\text{wor})} < 2^n$ if and only if $|\epsilon| < s_1\gamma$. The above is summarised as follows:

Case $|\epsilon| < \min(s_1, s_1\gamma)$**:**

- $P_S^{(\text{wor,adj})}$ is a decreasing function of N for $0 < N < N_0^{(\text{wor})}$;
- $P_S^{(\text{wor,adj})}$ is an increasing function of N for $N_0^{(\text{wor})} < N < N_1^{(\text{wor})}$;
- $P_S^{(\text{wor,adj})}$ is a decreasing function of N for $N_1^{(\text{wor})} < N \leq 2^n$;
- P_S attains a minima at $N_0^{(\text{wor})}$ and a maxima at $N_1^{(\text{wor})}$.

Case $s_1 < |\epsilon| < s_1\gamma$**:**

- $P_S^{(\text{wor,adj})}$ is an increasing function of N for $0 < N < N_1^{(\text{wor})}$;
- $P_S^{(\text{wor,adj})}$ is a decreasing function of N for $N_1^{(\text{wor})} < N \leq 2^n$;
- P_S attains a maxima at $N_1^{(\text{wor})}$.

Case $\max(s_1, s_1\gamma) < |\epsilon|$**:** $P_S^{(\text{wor,adj})}$ is an increasing function of N for $0 < N \leq 2^n$.

We have $s_1 = 2^{-2-n/2}$ and $\gamma = \Phi^{-1}\left(1 - 2^{m-a-1}/(2^m - 1)\right)$, where $2^m/(2^m - 1) < a \leq m$. The maximum value of γ is achieved for $a = m$ and this value is $\Phi^{-1}\left((2^{m+1} - 3)/(2^{m+1} - 2)\right)$ which is around 8.21 for $m \leq 64$. So, $s_1\gamma$ is not much greater than s_1. It seems reasonable to assume that in practice the value of ϵ will turn out to be such that $\max(s_1, s_1\gamma) < |\epsilon|$. Under this condition, both $P_S^{(\text{wr,adj})}$ and $P_S^{(\text{wor,adj})}$ are increasing functions of N for $0 < N \leq 2^n$. In other words, the anomalous non-monotonic behaviour will mostly not occur in practice. The non-monotonic behaviour is observed only when the value of $|\epsilon|$ is small enough to be less than either s_1 or $s_1\gamma$.

We further note the following point. The distribution of W_κ for $\kappa \neq \kappa^*$ is approximated as $\mathcal{N}(0, 2^{-n-2} + 1/(4N))$ for sampling with replacement and is approximated as $\mathcal{N}(0, 1/(4N))$ for sampling without replacement. As explained in Sections 5.1 and 5.2, both of these approximations require making the assumption that $(\mathfrak{p} - 1/2)^2$ is negligible for $\mathfrak{p} \in [1/2 - \vartheta_1, 1/2 + \vartheta_1]$. From (23), the assumption is meaningful only if we consider $s_1^2 2^{-n/2} = 2^{-2-n/2}$ to be negligible. So, the derivation of the distribution of W_κ for $\kappa \neq \kappa^*$ is meaningful only if $2^{-2-n/2}$ is considered to be negligible. Consequently, it is perhaps not meaningful to apply the analysis for values of $|\epsilon|$ lower than $2^{-2-n/2}$. This is a further argument that the analysis actually shows P_S is a monotonic increasing function of N in the range where the analysis is actually meaningful.

Remarks: The following comments are based on the assumption that $\gamma \approx \Phi^{-1}(1 - 2^{-a-1})$, i.e., $2^m \approx 2^m - 1$.

1. In [8] it was stated without proof that the first term of $P_S^{(\text{wr,adj})}$ given by (44) attains a maximum for $N = N_1^{(\text{wr})}$.

2. It was shown in [1] that the derivative of the first term of $P_S^{(\text{wor,adj})}$ given by (45) is zero for $N = N_1^{(\text{wor})}$ from which it was concluded without any further argument that $P_S^{(\text{wor,adj})}$ achieves a maxima at $N = N_1^{(\text{wor})}$.

We note that the complete picture of the dependence of the success probability on N was not provided in either [8] or [1].

7 Conclusion

In this paper, we have carried out a detailed and complete analysis of success probability of linear cryptanalysis. This has been done under a single unifying framework which provides a deeper insight and a better understanding of how the success probability behaves with respect to the data complexity.

References

[1] Tomer Ashur, Tim Beyne, and Vincent Rijmen. Revisiting the wrong-key-randomization hypothesis. *IACR Cryptology ePrint Archive*, 2016:990, 2016.

[2] Thomas Baignères, Pascal Junod, and Serge Vaudenay. How Far Can We Go Beyond Linear Cryptanalysis? In *Advances in Cryptology–ASIACRYPT 2004*, pages 432–450. Springer, 2004.

[3] Thomas Baignères, Pouyan Sepehrdad, and Serge Vaudenay. Distinguishing Distributions Using Chernoff Information. In *Provable Security*, pages 144–165. Springer, 2010.

[4] Alex Biryukov, Christophe De Cannière, and Michaël Quisquater. On Multiple Linear Approximations. In *Advances in Cryptology–CRYPTO 2004*, pages 1–22. Springer, 2004.

[5] Céline Blondeau, Benoît Gérard, and Kaisa Nyberg. Multiple Differential Cryptanalysis using LLR and χ^2 Statistics. In *Security and Cryptography for Networks*, pages 343–360. Springer, 2012.

[6] Céline Blondeau and Kaisa Nyberg. Joint data and key distribution of simple, multiple, and multidimensional linear cryptanalysis test statistic and its impact to data complexity. *Des. Codes Cryptography*, 82(1-2):319–349, 2017.

[7] Andrey Bogdanov and Elmar Tischhauser. On the wrong key randomisation and key equivalence hypotheses in matsui's algorithm 2. In *Fast Software Encryption - 20th International Workshop, FSE 2013, Singapore, March 11-13, 2013. Revised Selected Papers*, pages 19–38, 2013. `http://dx.doi.org/10.1007/978-3-662-43933-3_2`.

[8] Andrey Bogdanov and Elmar Tischhauser. On the Wrong Key Randomisation and Key Equivalence Hypotheses in Matsui's Algorithm 2. In *Fast Software Encryption*, pages 19–38. Springer, 2014.

[9] Joan Daemen and Vincent Rijmen. Probability Distributions of Correlation and Differentials in Block Ciphers. *Journal of Mathematical Cryptology JMC*, 1(3):221–242, 2007.

[10] Willliam Feller. *An Introduction to Probability Theory and Its Applications, Volume 1.* John Wiley & Sons, 2008.

[11] Benoît Gérard and Jean-Pierre Tillich. On linear cryptanalysis with many linear approximations. In *IMA International Conference on Cryptography and Coding*, pages 112–132. Springer, 2009.

[12] Carlo Harpes, Gerhard G. Kramer, and James L. Massey. A Generalization of Linear Cryptanalysis and the Applicability of Matsui's Piling-Up Lemma. In *Advances in Cryptology - EUROCRYPT '95, International Conference on the Theory and Application of Cryptographic Techniques, Saint-Malo, France, May 21-25, 1995, Proceeding*, pages 24–38, 1995. `http://link.springer.de/link/service/series/0558/bibs/0921/09210024.htm`.

[13] Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Multidimensional Extension of Matsui's Algorithm 2. In *Fast Software Encryption*, pages 209–227. Springer, 2009.

[14] Jialin Huang, Serge Vaudenay, Xuejia Lai, and Kaisa Nyberg. Capacity and Data Complexity in Multidimensional Linear Attack. In *Advances in Cryptology - CRYPTO 2015 - 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part I*, pages 141–160, 2015. `http://dx.doi.org/10.1007/978-3-662-47989-6_7`.

[15] Pascal Junod. On the Complexity of Matsui's Attack. In *Selected Areas in Cryptography*, pages 199–211. Springer, 2001.

[16] Pascal Junod. On the Optimality of Linear, Differential, and Sequential Distinguishers. In *Advances in Cryptology–EUROCRYPT 2003*, pages 17–32. Springer, 2003.

[17] Pascal Junod and Serge Vaudenay. Optimal Key Ranking Procedures in a Statistical Cryptanalysis. In *Fast Software Encryption*, pages 235–246. Springer, 2003.

[18] Burton S Kaliski Jr and Matthew JB Robshaw. Linear Cryptanalysis Using Multiple Approximations. In *Advances in Cryptology–Crypto'94*, pages 26–39. Springer, 1994.

[19] S. N. Lahiri and A. Chatterjee. A Berry-Esseen theorem for hypergeometric probabilities under minimal conditions. *Proceedings of the American Mathematical Society*, 135(5):1535–1545, 2007.

[20] Mitsuru Matsui. Linear Cryptanalysis Method for DES Cipher. In *Advances in Cryptology–EUROCRYPT'93*, pages 386–397. Springer, 1993.

[21] Mitsuru Matsui. The First Experimental Cryptanalysis of the Data Encryption Standard. In Y. G. Desmedt, editor, *Advances in Cryptology–Crypto'94*, pages 1–11. Springer, 1994.

[22] Sean Murphy. The Independence of Linear Approximations in Symmetric Cryptanalysis. *IEEE Transactions on Information Theory*, 52(12):5510–5518, 2006.

[23] Subhabrata Samajder and Palash Sarkar. Rigorous upper bounds on data complexities of block cipher cryptanalysis. *IACR Cryptology ePrint Archive*, 2015:916, 2015.

[24] Subhabrata Samajder and Palash Sarkar. Another look at normal approximations in cryptanalysis. *J. Mathematical Cryptology*, 10(2):69–99, 2016.

[25] Subhabrata Samajder and Palash Sarkar. A new test statistic for key recovery attacks using multiple linear approximations. *IACR Cryptology ePrint Archive*, 2016:404, 2016. To appear in the Proceedings of Mycrypt 2016.

[26] Ali Aydın Selçuk. On Probability of Success in Linear and Differential Cryptanalysis. *Journal of Cryptology*, 21(1):131–147, 2008.

[27] Elmar Tischhauser. Mathematical aspects of symmetric-key cryptography, 2012. PhD thesis.

[28] A. M. Walker. A Note on the Asymptotic Distribution of Sample Quantiles. *Journal of the Royal Statistical Society. Series B (Methodological)*, 30(3):pp. 570–575, 1968.

A Some Results on Statistics

A.1 Order Statistics

Selçuk [26] used a result on order statistics to derive an expression for the success probability. We briefly summarise this result.

Let $T_1, T_2, \ldots, T_{2^m-1}$ be independent and identically distribution random variables with common density function $f(x)$ and common distribution function $F(x)$. Let $T_{(1)}, T_{(2)}, \ldots, T_{(2^m-1)}$ be the random variables $T_1, T_2, \ldots, T_{2^m-1}$ sorted in ascending order. For $1 \leq a \leq 2^m - 1$, let $q = 1 - 2^{-a}$. Then the distribution of $T_{(2^m q)}$ approximately follows $\mathcal{N}(\mu_q, \sigma_q^2)$ where $\mu_q = F^{-1}(q)$ and $\sigma_q = 2^{-(m+a)/2}/f(\mu_q)$. This follows from a standard result in mathematical statistics. (See [28] for a proof of the asymptotic version of the result and [24] for a proof of the concrete version of the result.)

Further suppose $T_i = |W_i|$ where W_i follows $\mathcal{N}(0, \sigma_1)$. Then T_i follows a half-normal distribution whose density function is $f(y) = 2/(\sigma_1\sqrt{2\pi})\exp(-y^2/(2\sigma_1^2))$ and the distribution function $F(y)$ is obtained by integrating the density function $f(y)$. In this case, $T_{(2^m q)}$ approximately follows $\mathcal{N}(\mu_q, \sigma_q^2)$ where

$$\begin{aligned} \mu_q &= F^{-1}(q) = \sigma_1 \Phi^{-1}(q) = \sigma_1 \Phi^{-1}(1 - 2^{-a-1}); \\ \sigma_q &= \frac{1}{f(\mu_q)} 2^{-(m+a)/2} = \frac{\sigma_1}{2\phi\left(\Phi^{-1}\left(1 - 2^{-a-1}\right)\right)} 2^{-(m+a)/2}. \end{aligned}$$

A.2 Compound Normal

Recall that the density function of $\mathcal{N}(\mu, \sigma^2)$ is denoted as $\mathfrak{f}(x; \mu, \sigma^2)$.

Proposition 1.

$$\int_{-\infty}^{\infty} \mathfrak{f}(x; ay, \sigma_1^2) \cdot \mathfrak{f}(y; \mu, \sigma_2^2)\, dy \;=\; \mathfrak{f}(x; a\mu, \sigma_1^2 + a^2\sigma_2^2).$$

Proof.

$$\begin{aligned} &\mathfrak{f}(x; ay, \sigma_1^2) \cdot \mathfrak{f}(y; \mu, \sigma_2^2) \\ &= \left[\frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{(x-ay)^2}{2\sigma_1^2}\right\}\right] \cdot \left[\frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{(y-\mu)^2}{2\sigma_2^2}\right\}\right] \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{\sigma_1\sigma_2} \exp\left\{-\left(\frac{(x-ay)^2}{2\sigma_1^2} + \frac{(y-\mu)^2}{2\sigma_2^2}\right)\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{\sigma_1\sigma_2} \exp\left\{-\frac{1}{2\sigma_1^2\sigma_2^2}\left(\sigma_2^2x^2 + (\sigma_1^2 + a^2\sigma_2^2)y^2 - 2y(\sigma_2^2 ax + \sigma_1^2\mu) + \sigma_1^2\mu^2\right)\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{\sigma_1\sigma_2} \exp\left\{-\frac{\sigma_1^2 + a^2\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(\frac{\sigma_2^2x^2}{\sigma_1^2 + a^2\sigma_2^2} + y^2 - 2y\left(\frac{\sigma_2^2 ax + \sigma_1^2\mu}{\sigma_1^2 + a^2\sigma_2^2}\right) + \frac{\sigma_1^2\mu^2}{\sigma_1^2 + a^2\sigma_2^2}\right)\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{\sigma_1\sigma_2} \exp\left\{-\frac{\sigma_1^2 + a^2\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(\left(y - \frac{\sigma_2^2 ax + \sigma_1^2\mu}{\sigma_1^2 + a^2\sigma_2^2}\right)^2 + \frac{\sigma_2^2x^2 + \sigma_1^2\mu^2}{\sigma_1^2 + a^2\sigma_2^2} - \left(\frac{\sigma_2^2 ax + \sigma_1^2\mu}{\sigma_1^2 + a^2\sigma_2^2}\right)^2\right)\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{\sigma_1\sigma_2} \exp\left\{-\frac{\sigma_1^2 + a^2\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(y - \frac{\sigma_2^2 ax + \sigma_1^2\mu}{\sigma_1^2 + \sigma_2^2}\right)^2 - \frac{(x-a\mu)^2}{2(\sigma_1^2 + a^2\sigma_2^2)}\right\}. \end{aligned}$$

Therefore,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \mathfrak{f}(x; ay, \sigma_1^2) \cdot \mathfrak{f}(y; \mu, \sigma_2^2)\ dy \\
&= \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{\sigma_1\sigma_2} \exp\left\{-\frac{\sigma_1^2+a^2\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(y - \frac{\sigma_2^2 ax + \sigma_1^2\mu}{\sigma_1^2+\sigma_2^2}\right)^2 - \frac{(x-a\mu)^2}{2(\sigma_1^2+a^2\sigma_2^2)}\right\}\ dy \\
&= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-a\mu)^2}{2(\sigma_1^2+a^2\sigma_2^2)}\right\} \cdot \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left\{-\frac{\sigma_1^2+a^2\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(y - \frac{\sigma_2^2 ax + \sigma_1^2\mu}{\sigma_1^2+\sigma_2^2}\right)^2\right\}\ dy \\
&= \frac{1}{\sqrt{2\pi(\sigma_1^2+a^2\sigma_2^2)}} \exp\left\{-\frac{(x-a\mu)^2}{2(\sigma_1^2+a^2\sigma_2^2)}\right\} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(y - \sqrt{\frac{\sigma_1^2+a^2\sigma_2^2}{\sigma_1^2\sigma_2^2}}\frac{\sigma_2^2 ax + \sigma_1^2\mu}{\sigma_1^2+\sigma_2^2}\right)^2\right\}\ dy \\
&= \frac{1}{\sqrt{2\pi(\sigma_1^2+a^2\sigma_2^2)}} \exp\left\{-\frac{(x-a\mu)^2}{2(\sigma_1^2+a^2\sigma_2^2)}\right\} \\
&= \mathfrak{f}(x; a\mu, \sigma_1^2+a^2\sigma_2^2).
\end{aligned}$$

□

Proposition 2. *Let X and Y be two random variables such that $X \sim \mathcal{N}(aY, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu, \sigma_2^2)$, where $\mathfrak{a}$ is a constant. Then,*

$$X \sim \mathcal{N}(\mathfrak{a}\mu, \sigma_1^2 + \mathfrak{a}^2\sigma_2^2).$$

Proof. Let, $f_{X|Y}(x,y), f_{X,Y}(x,y)$ denote the conditional and joint distributions of the random variables X and Y, respectively. Also, let $f_Y(y)$ and $f_X(x)$ denote the marginal distributions of the random variables Y and X, respectively. Then,

$$f_{X|Y}(x,y) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{(x-\mathfrak{a}y)^2}{2\sigma_1^2}\right\} \text{ and } f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{(y-\mu)^2}{2\sigma_2^2}\right\}.$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy = \int_{-\infty}^{\infty} f_{X|Y}(x,y) f_Y(y)\ dy = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\sigma_1^2+\mathfrak{a}^2\sigma_2^2}} \cdot \exp\left\{-\frac{(x-\mathfrak{a}\mu)^2}{2(\sigma_1^2+\mathfrak{a}^2\sigma_2^2)}\right\}.$$

The last equality follows from Proposition 1. So, $X \sim \mathcal{N}(\mathfrak{a}\mu, \sigma_1^2+\mathfrak{a}^2\sigma_2^2)$. □

A.3 Hypergeometric Distribution

Suppose an urn contains $\mathfrak{N}$ distinguishable balls out of which $\mathfrak{R}$ are red and the rest are white. A sample of size $\mathfrak{n}$ is chosen from the urn without replacement. For $\mathfrak{k} \in \{0, \ldots, \mathfrak{n}\}$, the probability that there are exactly $\mathfrak{k}$ red balls in the sample is

$$\mathfrak{p}(\mathfrak{k}; \mathfrak{n}, \mathfrak{N}, \mathfrak{R}) = \frac{\binom{\mathfrak{R}}{\mathfrak{k}}\binom{\mathfrak{N}-\mathfrak{R}}{\mathfrak{n}-\mathfrak{k}}}{\binom{\mathfrak{N}}{\mathfrak{n}}}. \tag{48}$$

Here $\mathfrak{p}(\mathfrak{k}; \mathfrak{n}, \mathfrak{N}, \mathfrak{R})$ is the probability mass function of the hypergeometric distribution $\mathsf{H}(\mathfrak{k}; \mathfrak{n}, \mathfrak{N}, \mathfrak{R})$.

Let $p = \mathfrak{R}/\mathfrak{N}$ and $q = 1-p$. According to Problem 2 in Section 11 of Chapter II of Feller [10],

$$\binom{\mathfrak{n}}{\mathfrak{k}}\left(p - \frac{\mathfrak{k}}{\mathfrak{N}}\right)\left(q - \frac{\mathfrak{n}-\mathfrak{k}}{\mathfrak{N}}\right)^{\mathfrak{n}-\mathfrak{k}} < \mathfrak{p}(\mathfrak{k}; \mathfrak{n}, \mathfrak{N}, \mathfrak{R}) < \binom{\mathfrak{n}}{\mathfrak{k}} p^{\mathfrak{k}} q^{\mathfrak{n}-\mathfrak{k}} \left(1 - \frac{\mathfrak{n}}{\mathfrak{N}}\right)^{-\mathfrak{n}}. \tag{49}$$

Consequently, if $\mathfrak{N} \gg \mathfrak{n}$, then $\mathfrak{p}(\mathfrak{k}; \mathfrak{n}, \mathfrak{N}, \mathfrak{R}) \approx \binom{\mathfrak{n}}{\mathfrak{k}} p^{\mathfrak{k}} q^{\mathfrak{n}-\mathfrak{k}}$. In other words, if $\mathfrak{N} \gg \mathfrak{n}$, then the hypergeometric distribution is well approximated by the binomial distribution.

Another approximation of the hypergeometric distribution by the normal distribution appears in Problem 10 in Section 7 of Chapter VII of Feller [10]. Suppose $\mathfrak{t} \in (0,1)$ and p are such that $\frac{\mathfrak{n}}{\mathfrak{N}} \to \mathfrak{t}$, $\frac{\mathfrak{R}}{\mathfrak{N}} \to p$ as $\mathfrak{n}, \mathfrak{N}, \mathfrak{R} \to \infty$. Let $h = 1/\sqrt{\mathfrak{N}p(1-p)\mathfrak{t}(1-\mathfrak{t})}$ be such that $h(\mathfrak{k} - \mathfrak{n}p) \to x$. Then $\mathfrak{p}(\mathfrak{k}; \mathfrak{n}, \mathfrak{N}, \mathfrak{R}) \sim h\Phi(x)$. Consequently, if Y is a random variable following the hypergeometric distribution $\mathsf{H}(\mathfrak{k}; \mathfrak{n}, \mathfrak{N}, \mathfrak{R})$ then Y approximately follows $\mathcal{N}(p\mathfrak{n}, \mathfrak{N}p(1-p)\mathfrak{t}(1-\mathfrak{t}))$. Conditions for the normal approximation to be meaningful and bounds on the error in the approximation have been provided in [19]. Using $\mathfrak{n} = \mathfrak{N}\mathfrak{t}$, the random variable Y approximately follows $\mathcal{N}(p\mathfrak{n}, \mathfrak{n}p(1-p)(1-\mathfrak{t}))$.