

Research Article

Subhabrata Samajder and Palash Sarkar*

Another look at normal approximations in cryptanalysis

DOI: 10.1515/jmc-2016-0006

Received January 30, 2016; revised April 27, 2016; accepted April 27, 2016

Abstract: Statistical analysis of attacks on symmetric ciphers often requires assuming the normal behaviour of a test statistic. Typically such an assumption is made in an asymptotic sense. In this work, we consider concrete versions of some important normal approximations that have been made in the literature. To do this, we use the Berry–Esséen theorem to derive explicit bounds on the approximation errors. A basic mathematical requirement is that such approximation errors should be within reasonable bounds, a point which appears to have been overlooked in many of the earlier works on statistical aspects of cryptanalysis. Interpreting the error bounds in the cryptanalytic context yields several surprising results. One important implication is that this puts in doubt the applicability of the order statistics based approach for analysing key recovery attacks on block ciphers. This approach has been earlier used to obtain several results on the data complexities of (multiple) linear and differential cryptanalysis. The non-applicability of the order statistics based approach puts a question mark on the data complexities obtained using this approach. Fortunately, we are able to recover all of these results by utilising the hypothesis testing framework. This, however, necessitates using normal approximations for the χ^2 and the LLR test statistics considered in earlier works. These approximations themselves have issues which seem to be difficult to resolve satisfactorily. More generally, the message of our work is that all cryptanalytic attacks should properly derive and interpret the error bounds for any (normal) approximation that is made.

Keywords: Block cipher, linear/differential cryptanalysis, order statistics, normal distribution, Berry–Esséen theorem

MSC 2010: 94A60, 11T71, 68P25, 62P99

Communicated by: Alfred Menezes

1 Introduction

Let $X_1, \dots, X_\lambda$ be independent and identically distributed random variables having mean μ and variance σ^2 and taking values from a finite non-empty set. Asymptotically, $((X_1 + \dots + X_\lambda)/\lambda - \mu)/(\sigma/\sqrt{\lambda})$ follows the standard normal distribution as $\lambda \rightarrow \infty$. This is a consequence of the central limit theorem (CLT) and this asymptotic result is commonly assumed in many papers on statistical analysis of attacks on symmetric ciphers.

For example, if the independently and identically distributed random variables X_i follow Bernoulli(p) distribution, then it is well known that the normal approximation is “good” if p is close to $1/2$ and that it is “bad” if p is away from $1/2$. For the typical situation of (single) linear cryptanalysis, p is close to $1/2$ and so the normal approximation is “good”; while for the typical situation of (single) differential cryptanalysis, p is away from $1/2$ and so the normal approximation is “bad”. These observations are mentioned in several

Subhabrata Samajder: Applied Statistics Unit, Indian Statistical Institute, B. T. Road 203, Kolkata 700108, India, e-mail: subhabrata_r@isical.ac.in

***Corresponding author: Palash Sarkar:** Applied Statistics Unit, Indian Statistical Institute, B. T. Road 203, Kolkata 700108, India, e-mail: palash@isical.ac.in

papers in the literature (see, e.g., [12, p. 9]) which usually do not consider the error in approximation. This basic scenario itself points to a somewhat counter-intuitive aspect even for linear cryptanalysis. For a weak cipher the value of p will be further away from $1/2$ than for a strong cipher, but the statistical analysis will be able to say less about the weak cipher than about the strong cipher.

While the CLT is a basic result, there are other more complex situations in statistical cryptanalysis where normal approximations are assumed. Almost invariably, such assumptions are justified by references to a book or a paper in the area of mathematical statistics. On checking such references, one finds proofs which show the statements to hold asymptotically. The error in approximation is mostly not discussed in details or, even if it is mentioned, it is only considered in an asymptotic sense.

In cryptographic contexts, it is important to assess the error in the normal approximation for finite λ . It should be possible to bound the error and analyse conditions on λ for which the approximation is meaningful. The Berry–Essén theorem is the right tool for doing this. Such an approach can be called a concrete analysis as opposed to an asymptotic analysis. We note that in another branch of cryptography, namely provable security, the distinction between asymptotic analysis and concrete analysis is considered to be important.

Our contributions. The goal of this work is to take a deep look at some of the normality assumptions that have been made in the literature on block cipher cryptanalysis. Needless to say, considering all such works is beyond the scope of a single paper. Nonetheless, we do consider some of the key techniques. These include results on order statistics and the χ^2 and the log-likelihood ratio (LLR) based test statistics. Using the Berry–Essén theorem, we obtain bounds on the error in normal approximations for these situations. For the normal approximations to be meaningful, the approximation errors should be within reasonable bounds. We interpret the error bounds in the cryptanalytic contexts, i.e., we consider values of the attack parameters which ensure that the approximation errors are not too much. Doing this yields some very unexpected and counter-intuitive facts which invalidate several previous works. In some cases, we are able to recover the results following a different technique while in other cases, providing satisfactory replacements seems to be difficult.

For the convenience of describing our results, we set down some notation for the central quantities in cryptanalytic attacks on block ciphers that will be required in our work. For linear cryptanalysis, prior analysis of the block cipher will provide $\ell \geq 1$ linear approximations. If $\ell = 1$, the attack will be called single linear cryptanalysis or simply linear cryptanalysis, while if $\ell > 1$, the attack will be called multiple linear cryptanalysis where we will work with the joint distribution of the linear approximations. Similarly, for (multiple) differential cryptanalysis, the number of differentials will be denoted by $\nu \geq 1$ where the case of $\nu = 1$ denotes single differential cryptanalysis and the case of $\nu > 1$ denotes multiple differential cryptanalysis.

The goal of a cryptanalytic attack is to recover the correct value of a target sub-key consisting of m bits. Typically, an attack will provide a list of candidate values of the target sub-key. The attack is said to have advantage a ($1 \leq a \leq m$) if the (expected) number of candidate values is 2^{m-a} . The success probability of an attack will be denoted by P_S and is the probability that the correct value of the target sub-key is in the list of candidate values produced by the attack. To mount an attack, a set of plaintext-ciphertext pairs is required. The number of such pairs will be denoted by N and this number is called the data complexity of the attack. The main goal of a statistical analysis of cryptanalytic attack is to be able to express N in terms of a and P_S and also possibly m and ℓ or ν . This gives an idea of the number of plaintext-ciphertext pairs required to successfully mount an attack.

Below, we present a summary of our results.

Critique of the order statistics based approach to key recovery attacks. Selçuk [34] proposed an approach to analyse key recovery attacks based on asymptotic normal behaviour of order statistics of a sequence of random variables. We derive the concrete version of the normal approximation of the order statistics result. Using this, we are able to essentially invalidate the order statistics based approach for analysing key recovery attacks. In particular, if issues of convergence are taken seriously, then the order statistics based approach cannot be applied if either m is small or $m - a$ is small. So, if the attack targets a byte, then it becomes hard to justify the application of the order statistics based approach. Even more intriguing is the dependence on $m - a$ with m fixed. An attack becomes sharper as the value of a grows. The error analysis shows that the deviation

from normal grows as the value of a grows. So, the applicability of the order statistics based approach to attacks with advantage close to m is not clear.

Selçuk utilised the order statistics based approach to obtain expressions for the data complexities of single linear and differential attacks. This approach was later followed by Hermelin, Cho and Nyberg [20] for multiple linear cryptanalysis and by Blondeau, Gérard and Nyberg [11] for multiple differential cryptanalysis. Both these works obtained expressions for data complexities. The criticisms of the order statistics based approach puts into doubt the applicability of the data complexity results from [11, 20, 34].

Hypothesis testing based framework for key recovery attacks. This framework appears in the literature in the context of analysing distinguishing attacks [2, 3, 23]. In the context of key recovery attacks, it appears that the hypothesis testing framework has been less used. An early work which formulated this approach is [32]. A later work using the hypothesis testing framework is [12] though the main analysis in the paper was done using the order statistics framework. In fact, the important works [11, 20, 34] do not follow the hypothesis testing framework. We find this to be surprising, since the hypothesis testing framework is simpler and avoids the restrictions of the order statistics based approach. One possible advantage of using the order statistics framework over the hypothesis testing framework is that the former approach allows to specify an absolute bound on the size of the false alarm list while the latter approach only permits a bound on the expected size of the false alarm list. This disadvantage, however, is outweighed by the fact that ensuring the appropriateness of the normal approximation inherently required in the order statistics framework essentially invalidates the approach.

We show that the data complexity results from [11, 20, 34] can be recovered by following the hypothesis testing framework for analysing key recovery attacks. In re-deriving the results of [11, 20], we make use of normal approximations of a χ^2 -based and an LLR-based test statistics used in these works which did not, however, consider issues related to the error in such approximations.

On the appropriateness of using the χ^2 and the LLR-based test statistics. The χ^2 based test statistics used in [11, 20] requires several approximations. Analysing the conditions under which such approximations are meaningful show some very unexpected results. We mention these as applied to multiple linear cryptanalysis. These also hold for multiple differential cryptanalysis, though to a lesser extent.

- (i) As ℓ grows, the value of N should also grow. Considering N to be a measure of the difficulty of launching an attack, this result is counter-intuitive, since having more linear approximations should help the attack rather than make it more difficult.
- (ii) The value of ℓ itself has to be relatively high. If only a few linear approximations are available, then the χ^2 -based test statistics is inapplicable.
- (iii) The joint distribution of the ℓ linear approximations must be close to the uniform distribution at every point. This is non-intuitive, since if the analysis of the block cipher uncovers a large deviation from uniform at a certain point, then the χ^2 -based analysis is no longer applicable.

To a certain extent, some of these issues were passingly mentioned in the literature. For example, in [20], Hermelin, Cho and Nyberg mention “that the normal approximation of the wrong-key distribution is valid only when” $\ell > 5$. The error in the approximation is not considered at all and no justification is provided for this assertion. To the best of our knowledge, no detailed treatment of the approximation errors and their implications appeared earlier. On the other hand, even though we are able to uncover these issues, the problem of satisfactorily addressing these issues remains open.

The LLR based test statistics used in [11, 20] also requires normal approximation. Analysis of this approximation provides a bound on the error in approximation. It is, however, difficult to obtain a closed form expression for this bound. This makes it difficult to assess the quality of the normal approximation of the LLR test statistics. Further, the mean and variance of the normal distribution are approximated in terms of the capacity of the probability distribution under the correct key hypothesis. This approximation requires the Taylor series expansion. Analysing the condition for the Taylor series approximation to hold shows that the data complexity obtained using such approximation may not be applicable in certain practical scenarios. We argue that the Taylor series expansion is unnecessary and an equally useful expression for the data

complexity can be obtained without this expansion. This avoids making the approximation required for the Taylor series expansion and widens the applicability of the result.

In view of the above, our work possibly raises more questions than it settles. Nevertheless, we believe that these questions are important to the understanding of statistical cryptanalysis, and we hope that they will be addressed in the future.

Previous works. Linear cryptanalysis for block ciphers was introduced by Matsui in [29] using a single linear approximation. An earlier work of Tardy-Corffdir and Gilbert [35] proposed an attack on FEAL which also considered linear approximations of the FEAL S-boxes. Matsui [30] showed how linear cryptanalysis can be improved when two linear approximations were available. Independently, Kaliski and Robshaw [25] also showed that the availability of several linear approximations having the same key mask but different plaintext and ciphertext mask leads to an attack with a lower data complexity. Both attacks [25, 30] considered the linear approximations to be independent. Analysis under the independence assumption of the linear approximations was later done in [8, 24]. Murphy [31] pointed out that the independence assumption may not be valid.

A series of papers [2, 3, 23] carried out a systematic investigation of multiple linear cryptanalysis where the linear approximations are not necessarily independent. The motivation of these works was to analyse and obtain optimal distinguishers to distinguish between two distributions. This was done using the framework of hypothesis testing. Several important techniques, including the log-likelihood ratio test, were successfully developed to build optimal distinguishers.

Treatment of key recovery attacks using multiple linear approximation without requiring any independence assumption on the linear approximations was carried out by Hermelin, Cho and Nyberg [20]. The work proposed two methods, one based on the χ^2 statistics and the other based on a log-likelihood ratio (LLR) test and computed expressions for the data complexity of the corresponding methods. This work used the key ranking framework initially proposed by Matsui [30] and employed the order statistics based approach outlined by Selçuk [34].

Differential cryptanalysis was proposed by Biham and Shamir [6] where a single differential was used. In a later work [7], Biham and Shamir improved the attack by considering together several differentials having the same output difference. Knudsen [26] introduced the idea of truncated differential attacks using multiple differentials with different output differences. The more general case of multiple differential attacks where both the input and output differences can be different was considered by Blondeau and Gérard in [10]. Later Blondeau, Gérard and Nyberg [11] considered multiple differential cryptanalysis with the same input difference but different output differences. They used the LLR and the χ^2 test statistic along with the ranking approach of Selçuk to obtain expressions for data complexities of the attacks.

A general study of data complexity and success probability of statistical attacks was carried out by Blondeau, Gérard and Tillich [12]. The paper clearly identified the two approaches of hypothesis testing and key ranking.

Many works in the literature are concerned with obtaining linear approximations and differential characteristics of practical block ciphers. These works do not affect the statistical theory behind the analysis and so are not relevant in the current context. There have been several works which analyse the key dependent behaviour of linear and differential characteristics [1, 9, 13, 14, 27]. These works also require approximations and the message of the present work should also be applicable to these works.

Outline of the paper. In Section 2, we state the Berry–Esséen theorem in the form that is useful for our purposes. Using the Berry–Esséen theorem, we derive an upper bound on the normal approximation of order statistics. Section 3 provides a sketch of the background on block cipher cryptanalysis with emphasis on linear cryptanalysis. The critique of the key recovery attacks using the order statistics based approach is given in Section 4. As mentioned earlier, this puts the approach into doubt. The alternative hypothesis testing based framework for key recovery attacks is outlined in Section 5. In Section 6, we re-derive the results on single and multiple linear cryptanalysis appearing in [20, 34] using the hypothesis testing based framework. Section 7 does the same for the results on single and multiple differential cryptanalysis appearing in [11, 34]. In re-deriving these results, we make use of the χ^2 and the LLR test statistics and associated approximations. In

Section 8, we analyse in depth the manner in which the χ^2 test statistics is used. In Section 9, we consider the error in approximation of the distribution of the LLR test statistics by the normal distribution. Finally, Section 10 concludes the paper.

2 Some results on statistics

In many situations it is required to approximate a scaled sum of random variables by the standard normal distribution. Typically, the central limit theorem is used to justify such approximation. The quality of approximation is explicitly given by the Berry–Essén theorem which is a standard result in probability theory [5, 16, 17]. We state the theorem in a form that is adequate for our purpose (cf. [18]).

Theorem 2.1. *Suppose $V_1, \dots, V_\lambda$ are independent and identically distributed random variables with $E[V_i] = 0$, $E[V_i^2] = \sigma^2$ and $E[|V_i|^3] = \rho < \infty$. Let $D_\lambda = (V_1 + \dots + V_\lambda)/(\sqrt{\lambda}\sigma)$ and let $\Phi(\cdot)$ be the distribution function of the standard normal distribution. Then there is a positive constant C such that for all real x and positive integer λ ,*

$$|\Pr[D_\lambda \leq x] - \Phi(x)| \leq \frac{C\rho}{\sigma^3 \sqrt{\lambda}}. \quad (2.1)$$

The best known upper and lower bounds for C are $C < 0.4748$ (see [36]) and $C \geq 0.40973$ (see [17]). Theorem 2.1 can be used to obtain a concrete estimate of the goodness of approximation of the distribution of the scaled sum of a sequence of Bernoulli distributed random variables by the standard normal distribution.

Corollary 2.2. *Let $Y_1, \dots, Y_\lambda$ be a sequence of independent Bernoulli(p) random variables with $p \in (0, 1)$. Let $c = 2(p - 1/2)$. Then for every real x and positive integer λ ,*

$$\left| \Pr\left[\frac{(Y_1 + \dots + Y_\lambda)/\lambda - p}{\sigma/\sqrt{\lambda}} < x\right] - \Phi(x) \right| \leq \frac{\Lambda_p}{\lambda^{1/2}}, \quad (2.2)$$

where

$$\Lambda_p = \frac{C \times (p^2 + (1-p)^2)}{(p(1-p))^{1/2}} = \frac{C \times (1+c^2)}{(1-c^2)^{1/2}}.$$

Consequently, if the bound on the right-hand side of (2.2) is at most $2^{-\varepsilon}$, then

$$\lambda \geq 2^{2\varepsilon} \Lambda_p^2 = 2^{2\varepsilon} C^2 \times \frac{(p^2 + (1-p)^2)^2}{p(1-p)} = 2^{2\varepsilon} C^2 \times \frac{(1+c^2)^2}{1-c^2}.$$

Remark 2.3. In (2.1) and (2.2) the difference between the two probabilities is given by an upper bound. Our arguments regarding tightness will be based on considering the upper bound to be less than some pre-specified quantity $2^{-\varepsilon}$. While this is a sufficient condition for the actual difference of probabilities to be less than $2^{-\varepsilon}$, it need not be necessary. Obtaining tighter bounds may change some of the analysis and observations that we make. We make two remarks regarding this aspect. First, one has to proceed with the currently best known bounds. A similar approach is taken in provable security where one analyses the security of a scheme with respect to the best known bounds. The second issue is that the Berry–Essén theorem has a long history and it is perhaps unlikely that tighter bounds can be (easily) obtained.

Type of approximations. Suppose it is desired that the difference of the two probabilities in (2.2) is at most $2^{-\varepsilon}$. This is ensured by requiring the upper bound in (2.2) to be at most $2^{-\varepsilon}$. Depending on the value of p , this gives rise to two types of conditions.

- Condition 1: Suppose $p \approx 1/2$. Then $c \approx 0$ and $\Lambda_p \approx C$ which is independent of p . For $\lambda \approx C^2 2^{2\varepsilon}$, the bound on the right-hand side of (2.2) is $\approx 2^{-\varepsilon}$. In this case, the value of λ for which the approximation is good is independent of c .
- Condition 2: Suppose p is close to 0. Then $1-p \approx 1$ and $p^2 + (1-p)^2 \approx 1$ and so $\Lambda_p \approx C/\sqrt{p}$. So, for $\lambda \approx C^2 2^{2\varepsilon}/p$, the bound on the right-hand side of (2.2) is $\approx 2^{-\varepsilon}$. Similarly, if p is close to 1, for $\lambda \approx C^2 2^{2\varepsilon}/(1-p)$, the bound on the right-hand side of (2.2) is $\approx 2^{-\varepsilon}$. In both cases, the value of λ for which the approximation is good depends on p .

Approximations under the first condition are easier to ensure while approximations under the second condition usually entail some (usually unstated) assumptions. Later we will see examples of approximations under both types of conditions.

2.1 Order statistics

Definition 2.4. Let $X_1, X_2, \dots, X_\lambda$ be independent and identically distributed random variables. Arrange the values of $X_1, X_2, \dots, X_\lambda$ in increasing order, resulting in $X_{(1)}, X_{(2)}, \dots, X_{(\lambda)}$. The statistic $X_{(i)}$ is called the i -th order statistic of the sample $X_1, X_2, \dots, X_\lambda$.

The following is a well-known result in the area of order statistics. We have reformulated the statement to provide an explicit upper bound. The proof that we provide is based on ideas from Walker [37] in combination with an application of the Berry–Esséen theorem. In the literature that we searched, we did not find the concrete version of the result and the proof that we provide below.

Theorem 2.5. Let $T_1, \dots, T_\lambda$ be independent and identically distributed random variables following a distribution function $F(x)$ which is continuous and strictly increasing for $0 < F(x) < 1$. Let $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(\lambda)}$ be the order statistics of $T_1, \dots, T_\lambda$. Let $q \in (0, 1)$ be a real number and $\xi_q = F^{-1}(q)$ and suppose that $F'(\xi_q) = f(\xi_q)$ exists and is positive. Let $\{r(\lambda)\}$ be a sequence of integers such that

$$\lim_{\lambda \rightarrow \infty} \lambda^{-1/2} (r(\lambda) - \lambda q) = 0.$$

Let $W_\lambda = \lambda^{1/2} (T_{(r(\lambda))} - \xi_q)$. Then for each real number z ,

$$\left| \Pr[W_\lambda \leq z] - \Phi\left(\frac{zf(\xi_q)}{\sigma_\lambda} + \frac{(\xi_\lambda - \lambda^{-1/2}(r(\lambda) - \lambda q))}{\sigma_\lambda}\right) \right| \leq \frac{\Lambda_{q_\lambda}}{\lambda^{1/2}}.$$

Here

$$q_\lambda = F\left(\xi_q + \frac{z}{\sqrt{\lambda}f(\xi_q)}\right), \quad \xi_\lambda = -z + \sqrt{\lambda}(q_\lambda - q), \quad \Lambda_{q_\lambda} = C \times \frac{(1 - q_\lambda)^2 + q_\lambda^2}{(q_\lambda(1 - q_\lambda))^{1/2}} \quad (2.3)$$

and $\Phi(\cdot)$ is the standard normal distribution function. Consequently, as $\lambda \rightarrow \infty$ the sequence $\{W_\lambda\}$ converges in distribution to $\mathcal{N}(0, q(1 - q)/f^2(\xi_q))$.

Proof. Define $Z_\lambda = f(\xi_q)W_\lambda = \lambda^{1/2}f(\xi_q)(T_{(r(\lambda))} - \xi_q)$ and note

$$\Pr[Z_\lambda \leq z] = \Pr\left[T_{(r(\lambda))} \leq \xi_q + \frac{z}{\lambda^{1/2}f(\xi_q)}\right]. \quad (2.4)$$

Define $U_i = F(T_i)$ for $i = 1, \dots, \lambda$. Since $T_1, \dots, T_\lambda$ are independent and identically distributed, the random variables $U_1, \dots, U_\lambda$ are also independent and identically distributed. Since $F(\cdot)$ is a distribution function, the random variables U_i take values in $(0, 1)$. Also, since $F(x)$ is strictly increasing, it is invertible. For any $a \in (0, 1)$,

$$\Pr[U_i \leq a] = \Pr[F(T_i) \leq a] = \Pr[T_i \leq F^{-1}(a)] = \int_{-\infty}^{F^{-1}(a)} f(x)dx = F(F^{-1}(a)) = a.$$

This shows that the random variables U_i are uniformly distributed in $(0, 1)$. Further, since $F(x)$ is strictly increasing, $T_i < T_j$ implies $U_i = F(T_i) < F(T_j) = U_j$. So, if π is a permutation of $\{1, \dots, n\}$ such that $T_{(i)} = T_{\pi(i)}$ then $U_{(i)} = U_{\pi(i)}$, i.e., the random variables U_i preserve the order of the random variables T_i and so $U_{(i)} = F(T_{(i)})$. Applying F to both sides of the inequality within the expression for probability on the right-hand side of (2.4), we obtain

$$\Pr[Z_\lambda \leq z] = \Pr[U_{(r(\lambda))} \leq q_\lambda].$$

Define binary-valued random variables $Y_1, \dots, Y_\lambda$ such that $Y_i = 1$ if $U_i \leq q_\lambda$ and $Y_i = 0$ otherwise; further define $A_\lambda = Y_1 + \dots + Y_\lambda = \#\{i : U_i \leq q_\lambda\}$. The variables Y_i are independent Bernoulli(q_λ) distributed. Let

$$\sigma_\lambda^2 = E[(Y_i - q_\lambda)^2] \quad \text{and} \quad \rho_\lambda = E[|Y_i - q_\lambda|^3].$$

The event $U_{(r(\lambda))} \leq q_\lambda$ represents the fact that there are at least $r(\lambda)$ of the random variables U_i which are less than q_λ . This is equivalent to the event $A_\lambda \geq r(\lambda)$. So,

$$\Pr[Z_\lambda \leq z] = \Pr[U_{(r(\lambda))} \leq q_\lambda] = \Pr[A_\lambda \geq r(\lambda)].$$

We now propose to use the Berry–Essén theorem as given in Corollary 2.2. Define

$$D_\lambda = \frac{(Y_1 + \cdots + Y_\lambda)/\lambda - q_\lambda}{\sigma/\sqrt{\lambda}}.$$

From Corollary 2.2, for every real x and positive integer λ ,

$$\frac{C \times \rho_\lambda}{\sigma_\lambda^3 \lambda^{1/2}} \geq |\Pr[D_\lambda < x] - \Phi(x)| = |\Pr[D_\lambda \geq x] - \Phi(-x)|. \quad (2.5)$$

We connect the bound to $\Pr[Z_\lambda \leq z]$. From the definition of D_λ it follows that $D_\lambda = \lambda^{-1/2}(A_\lambda - \lambda q_\lambda)/\sigma_\lambda$ and so

$$\Pr[Z_\lambda \leq z] = \Pr[A_\lambda \geq r(\lambda)] = \Pr\left[D_\lambda \geq \lambda^{-1/2} \frac{r(\lambda) - \lambda q_\lambda}{\sigma_\lambda}\right]. \quad (2.6)$$

We now take a more careful look at q_λ . Define

$$\zeta_\lambda = \frac{z}{f(\xi_q)} \left(\frac{F(\xi_q + \frac{z}{\lambda^{1/2}f(\xi_q)}) - q}{\frac{z}{\lambda^{1/2}f(\xi_q)}} - f(\xi_q) \right). \quad (2.7)$$

Set $\Delta x = z/(\lambda^{1/2}f(\xi_q))$ and note that Δx tends to 0 as λ tends to infinity. We have

$$\lim_{\lambda \rightarrow \infty} \zeta_\lambda = \lim_{\lambda \rightarrow \infty} \frac{z}{f(\xi_q)} \left(\frac{F(\xi_q + \Delta x) - q}{\Delta x} - f(\xi_q) \right) = \frac{z}{f(\xi_q)} \left(\lim_{\Delta x \rightarrow 0} \frac{F(\xi_q + \Delta x) - q}{\Delta x} - f(\xi_q) \right) = 0.$$

The last equality follows from the fact that the derivative of $F(x)$ at ξ_q exists and is equal to $f(\xi_q)$. From (2.7) we can write

$$q_\lambda = F\left(\xi_q + \frac{z}{\lambda^{1/2}f(\xi_q)}\right) = q + \frac{1}{\sqrt{\lambda}}(z + \zeta_\lambda). \quad (2.8)$$

So, q_λ tends to q as λ tends to infinity. Using the expression for q_λ given by (2.8) in (2.6), we obtain

$$\Pr[Z_\lambda \leq z] = \Pr\left[D_\lambda \geq \frac{\lambda^{-1/2}(r(\lambda) - \lambda q) - \zeta_\lambda}{\sigma_\lambda} - \frac{z}{\sigma_\lambda}\right].$$

Setting $x = \lambda^{-1/2}(r(\lambda) - \lambda q)/\sigma_\lambda - \zeta_\lambda/\sigma_\lambda - z/\sigma_\lambda$ and substituting in (2.5), we obtain

$$\left| \Pr[Z_\lambda \leq z] - \Phi\left(\frac{z}{\sigma_\lambda} + \frac{\zeta_\lambda - \lambda^{-1/2}(r(\lambda) - \lambda q)}{\sigma_\lambda}\right) \right| \leq \frac{C \times \rho_\lambda}{\sigma_\lambda^3 \lambda^{1/2}}. \quad (2.9)$$

Note that $W_\lambda = Z_\lambda/f(\xi_q)$. Then from (2.9) we obtain

$$\left| \Pr[W_\lambda \leq z] - \Phi\left(\frac{zf(\xi_q)}{\sigma_\lambda} + \frac{\zeta_\lambda - \lambda^{-1/2}(r(\lambda) - \lambda q)}{\sigma_\lambda}\right) \right| \leq \frac{C \times \rho_\lambda}{\sigma_\lambda^3 \lambda^{1/2}}.$$

This shows the desired bound.

For the last statement, it is sufficient to note that as $\lambda \rightarrow \infty$ the following hold: $\sigma_\lambda^2 \rightarrow q(1-q)$; $\zeta_\lambda \rightarrow 0$; $\lambda^{-1/2}(r(\lambda) - \lambda q) \rightarrow 0$ (from the given condition). $\square$

3 Background for cryptanalysis

We provide a brief description of the relevant details of a block cipher. Based on this, we discuss the background for linear cryptanalysis. The case of differential cryptanalysis is described later.

Iterated block cipher. A block cipher is a function $E : \{0, 1\}^k \times \{0, 1\}^n \rightarrow \{0, 1\}^n$ such that for each $K \in \{0, 1\}^k$, the function $E_K(\cdot) := E(K, \cdot)$ is a permutation of $\{0, 1\}^n$. The k -bit quantity K is called the secret key; the n -bit input to the block cipher is called the plaintext; and the n -bit output of the block cipher is called the ciphertext.

Most practical constructions of block ciphers are obtained by iterating one (or several) functions over several rounds. The secret key is expanded using a function to obtain round keys. The input to and the output of each round are n -bit strings and the round keys are also n -bit strings. Denote by $k^{(0)}, k^{(1)}, \dots$ the round keys, and by $R_{k^{(0)}}^{(0)}, R_{k^{(1)}}^{(1)}, \dots$ the round functions. The round functions are bijections of $\{0, 1\}^n$. Further, denote by $K^{(i)}$ the concatenation of the first i round keys, and by $E_{K^{(i)}}^{(i)}$ the composition of the first i round functions, i.e.,

$$E_{K^{(i)}}^{(i)} = R_{k^{(i-1)}}^{(i-1)} \circ \dots \circ R_{k^{(0)}}^{(0)} = R_{k^{(i-1)}}^{(i-1)} \circ E_{K^{(i-1)}}^{(i-1)}.$$

A block cipher may have many rounds and a cryptanalytic effort may target only a few of these rounds. Suppose that an attack targets $r + 1$ rounds. For a plaintext P , let C be the output after $r + 1$ rounds and B be the output after r rounds. So, $B = E_{K^{(r)}}^{(r)}(P)$ and $C = R_{k^{(r)}}^{(r)}(B)$.

Relations between plaintext and the input to the last round. Detailed analysis of the structure of a block cipher reveals one or more possible relations between a plaintext P ; the input to the last round B ; and possibly $K^{(r)}$, the concatenation of the first r round keys. Such relations can be in the form of a linear function or in the form of a differential as we explain later. In most cases, the relations hold with some probabilities. The probability is taken over uniform random choice of P . If there are more than one relations, then we may have to consider the joint distribution of the probabilities that these relations hold. Obtaining relations and their (possibly joint) distribution is a non-trivial task and often requires a great deal of experience and ingenuity. These relations form the basic platform on which a statistical analysis of an attack can be carried out.

Target sub-key. A single relation between P and B will involve only a subset of the bits of B . If several (or multiple) relations between P and B are known, it is required to consider the subset of the bits of B which cover all the relations. Obtaining these bits from C will require a partial decryption of the last round. Such a partial decryption will involve a subset of the bits of secret key (or of the last round key). Obtaining the correct values of these key bits is the goal of the attack and these bits will be called the target sub-key and sometimes to be the last round key bits. The number of such bits will be denoted by m . So, m key bits are sufficient to partially decrypt C to obtain the bits of B which are involved in any of the relation between P and B . There are 2^m possible choices of the target sub-key bits out of which one is correct and all others are incorrect. The requirement is to pick out the correct key. Let $M = 2^m - 1$ be the number of incorrect choices of the target sub-key.

Setting of an attack. Suppose there are N plaintext-ciphertext pairs (P_i, C_i) , $i = 1, \dots, N$ which have been generated using the correct key and are available. The plaintexts $P_1, \dots, P_N$ are supposed to be independent and uniformly distributed over the plaintext space. This is true for linear cryptanalysis but not for differential cryptanalysis which we discuss in more details later. For each choice κ of the last round key bits, it is possible to invert C_j to obtain the relevant bits of $B_{\kappa,j}$. The relevant bits are those which are required to evaluate the relations discovered in the prior analysis of the block cipher. Note that $B_{\kappa,j}$ depends on κ even though C_j may not. If κ is the correct choice for the target sub-key, then C_j indeed depends on κ , otherwise C_j has no relation to κ .

Given P_j and the relevant bits of $B_{\kappa,j}$ it is possible to evaluate all the known relations. From the result of these evaluations, one defines a test statistic T_κ . Since there are a total of 2^m possible values of κ , there are also 2^m random variables T_κ . These random variables are assumed to be independent and the distribution of these random variables depends on whether κ is correct or incorrect. It is also assumed that the distributions of T_κ for incorrect κ are identical. This ‘wrong key assumption’ was stated in [19]. As mentioned in Section 1, some recent works have considered the situation where the distributions of T_κ are dependent on κ . This setting was not used in the papers [11, 20, 34] and we also do not consider this setting. For an attack to be possible, it is required to obtain the two possible distributions of T_κ – one when κ is the correct choice and the other when κ is an incorrect choice.

3.1 Linear cryptanalysis

Assume that the analysis of the structure of the block cipher provides $\ell \geq 1$ linear approximations. These are given by masks $\Gamma_P^{(i)}$, $\Gamma_B^{(i)}$ and $\Gamma_K^{(i)}$ for $i = 1, \dots, \ell$. The subscript P denotes the plaintext mask, the subscript B denotes the mask after r rounds, and the subscript K denotes the mask for $K^{(r)}$. So, $\Gamma_P^{(i)}$ and $\Gamma_B^{(i)}$ are in $\{0, 1\}^n$ and $\Gamma_K^{(i)}$ is in $\{0, 1\}^{nr}$. If $\ell > 1$, then the attack is called multiple linear cryptanalysis, and if $\ell = 1$, we will call the attack single linear cryptanalysis, or simply, linear cryptanalysis. Define

$$L_i = \langle \Gamma_P^{(i)}, P \rangle \oplus \langle \Gamma_B^{(i)}, B \rangle \quad \text{for } i = 1, \dots, \ell.$$

Inner key bits. For a fixed but unknown key $K^{(r)}$, the quantity $z_i = \langle \Gamma_K^{(i)}, K^{(r)} \rangle$ is a single unknown bit. Denote by $z = (z_1, \dots, z_\ell)$ the collection of the ℓ bits arising in this manner. The key masks $\Gamma_K^{(1)}, \dots, \Gamma_K^{(\ell)}$ are known. So, z is determined only by the unknown key $K^{(r)}$. The bits represented by z are called the inner key bits. The key $K^{(r)}$ is unknown but fixed and so there is no randomness in $K^{(r)}$. Correspondingly, z is also unknown but fixed and there is no randomness in z .

Consider a uniform random choice of P . The round functions are deterministic bijections and so the uniform distribution on P induces a uniform distribution on B . Each L_i is a random variable which can take the values 0 or 1. The randomness of L_i arises solely from the randomness of P . Define the random variable X to be the following:

$$X = (L_1, \dots, L_\ell).$$

So, X is distributed over $\{0, 1\}^\ell$ and its distribution is determined by the distribution of the L_i which in turn is determined by the distribution of P .

A single linear approximation is of the form

$$L_i = \langle \Gamma_K^{(i)}, K^{(r)} \rangle = z_i. \quad (3.1)$$

Note that we do not assume any randomness over the key $K^{(r)}$ and so the bits z_i have no randomness even though they are unknown. So, the distribution of $L_i \oplus z_i$ is determined completely by the distribution of L_i .

Joint distribution parameterised by inner key bits. A linear approximation of the type given by (3.1) holds with some probability over the uniform random choice of P . The random variables $L_1, \dots, L_\ell$ are not necessarily independent. The joint distribution of these variables is given as follows: For $z = (z_1, \dots, z_\ell)$ and $\eta = (\eta_1, \dots, \eta_\ell) \in \{0, 1\}^\ell$, define

$$p_z(\eta) = \Pr[L_1 = \eta_1 \oplus z_1, \dots, L_\ell = \eta_\ell \oplus z_\ell] = \frac{1}{2^\ell} + \epsilon_\eta(z),$$

where $-1/2^\ell \leq \epsilon_\eta(z) \leq 1 - 1/2^\ell$.

The vector $\tilde{p}_z := (p_z(0), \dots, p_z(2^\ell - 1))$ is a probability distribution, where the integers $\{0, \dots, 2^\ell - 1\}$ are identified with the set $\{0, 1\}^\ell$. For each choice of z , we obtain a different distribution. These distributions are, however, related to each other. Suppose $z' = z \oplus \beta$ for some $\beta \in \{0, 1\}^\ell$. Then it is easy to verify that $\epsilon_\eta(z') = \epsilon_{\eta \oplus \beta}(z)$. It follows that

$$p_{z \oplus \beta}(\eta) = p_z(\eta \oplus \beta). \quad (3.2)$$

Let $\tilde{p}$ be the probability distribution $\tilde{p} := \tilde{p}_{0^\ell}$ so that

$$p(\eta) = 1/2^\ell + \epsilon_\eta \quad \text{for } \eta \in \{0, 1\}^\ell. \quad (3.3)$$

For the probability distribution $\tilde{p}$, define capacity $C(\tilde{p})$ as follows:

$$C(\tilde{p}) = 2^\ell \times \sum_{\eta \in \{0, 1\}^\ell} \epsilon_\eta^2.$$

Some notation. There are N plaintext-ciphertext pairs (P_j, C_j) for $j = 1, \dots, N$. For a choice κ of the target sub-key, the C_j are partially decrypted to obtain the relevant bits of $B_{\kappa, j}$. For $\kappa \in \{0, \dots, 2^m - 1\}$, $j = 1, \dots, N$

and $i = 1, \dots, \ell$, define

$$L_{\kappa,j,i} = \langle \Gamma_P^{(i)}, P_j \rangle \oplus \langle \Gamma_B^{(i)}, B_{\kappa,j} \rangle, \quad X_{\kappa,j} = (L_{\kappa,j,1}, \dots, L_{\kappa,j,\ell}).$$

Consider the values of $X_{\kappa,j}$ to be given by integers in the range 0 to $2^\ell - 1$. Define

$$Q_{\kappa,\eta} = \#\{j : X_{\kappa,j} = \eta\} \quad \text{for } \eta = 0, \dots, 2^\ell - 1. \quad (3.4)$$

So, for a fixed κ , the random variable $Q_{\kappa,\eta}$ is the number of occurrences of η in $X_{\kappa,1}, \dots, X_{\kappa,N}$.

4 Key recovery attack via order statistics: A critique

It is desirable to obtain an estimate of N , the number of plaintext-ciphertext pairs, that will be required for the attack to be successful. Selçuk [34] proposed a technique for doing this. The technique is generic and simple. The top level idea is the following.

Note that the test statistic T_κ is indexed by the m -bit string κ . For simplicity of notation, Selçuk considers these test statistics to be $T_0, \dots, T_{2^m-1}$ and assumes T_0 to be the test statistics corresponding to the correct choice of κ . Let $T_{(1)}, \dots, T_{(M)}$ be the order statistics of $T_1, \dots, T_M$. Recall $M = 2^m - 1$.

Advantage of an attack. Selçuk defines an attack to have an a -bit advantage if $T_0 > T_{(2^{m-a})}$, i.e., T_0 is ranked within the top 2^{m-a} values. In other words, an attack with a -bit advantage is said to be successful if $T_0 > T_{(2^{mq})}$ where $q = 1 - 2^{-a}$.

Selçuk's key idea was to use Theorem 2.5 to approximate the distribution of $T_{(2^{mq})}$. The statement of Theorem 2.5 considers the order statistics of λ random variables. In the current context, these λ variables are $T_1, \dots, T_M$. So, to apply Theorem 2.5 it is required to set

$$\lambda = M = 2^m - 1.$$

The sequence $r(\lambda)$ is defined to be $r(\lambda) = \lfloor q\lambda \rfloor + 1$ and so the convergence condition on $r(\lambda)$ holds. Also,

$$r(\lambda) = r(M) = \lfloor q(M) \rfloor + 1 = \lfloor q2^m - q \rfloor + 1 = q2^m - 1 + 1 = q2^m$$

and so $T_{(r(\lambda))} = T_{(q2^m)}$.

For "sufficiently large" M , the upper bound given by Theorem 2.5 is assumed to be negligible and also the expression for Φ is approximated by $\Phi(zf(\xi_q)/(q(1-q))^{1/2})$. This shows that the random variable $T_\lambda = \sqrt{\lambda}(T_{(q2^m)} - \xi_q)$ approximately follows $\mathcal{N}(0, q(1-q)/f^2(\xi_q))$. From this, it is argued that $T_{(2^{mq})}$ approximately follows $\mathcal{N}(\xi_q, q(1-q)/(\lambda f^2(\xi_q)))$. Using $\lambda = M - 1 = 2^m - 1$, the random variable $T_{(2^{mq})}$ approximately follows $\mathcal{N}(\xi_q, q(1-q)/((2^m - 1)f^2(\xi_q)))$.

We now consider several criticisms of this approach.

4.1 Normality requirement

In the order statistics based approach, the event success is defined to be $T_0 > T_{(2^{mq})}$. So, the probability of success is the probability that $T_0 - T_{(2^{mq})} > 0$. To compute this probability, the distribution of $T_0 - T_{(2^{mq})}$ is required. The result on order statistics shows that $T_{(2^{mq})}$ approximately follows a normal distribution. If it also turns out that T_0 follows a normal distribution, then $T_0 - T_{(2^{mq})}$ follows a normal distribution and it becomes possible to get an expression for the success probability. On the other hand, if T_0 does not follow the normal distribution, then in general it is not clear how to obtain the distribution for $T_0 - T_{(2^{mq})}$. To tackle this situation, a normal approximation of the distribution of T_0 is assumed and then the methodology is applied. Requiring both the distributions of the test statistics for correct and incorrect choices of the target sub-key to be normal is a restriction.

Blondeau, Gérard and Tillich [12] considered the scenario where the random variable T_0 follows a binomial distribution. The above approach was extended to compute success probability without approximating the distribution of T_0 by normal. While this is an interesting theoretical result, to use the expression for data complexity, it would be required to evaluate the binomial distribution which is computationally quite expensive.

4.2 Convergence and tightness issues

Considering convergence and tightness issues gives rise to two constraints.

Constraint 1: m must be large. In the proof of Theorem 2.5, for the convergence in distribution to $\mathcal{N}(0, q(1-q)/f^2(\xi_q))$ it is required that $q_\lambda \rightarrow q$ as $\lambda \rightarrow \infty$. This raises the question of how large λ should be for the convergence to be meaningful, and further whether and what condition does this give rise to.

Note that

$$q_\lambda = F\left(\xi_q + \frac{z}{\sqrt{\lambda}f(\xi_q)}\right) = q + \frac{1}{\sqrt{\lambda}}(z + \zeta_\lambda)$$

from which we get the condition that $q_\lambda \rightarrow q$ as $\lambda \rightarrow \infty$. The rate of convergence of q_λ to q is determined by $\lambda^{-1/2}$. So, for a fixed z , if we wish $|q_\lambda - q|$ to be less than $2^{-\varepsilon}$, then λ has to be proportional to $2^{2\varepsilon}$. Recalling that $\lambda = M = 2^m - 1$, this shows that $m \approx 2\varepsilon$ for the error in approximation to be about $2^{-\varepsilon}$.

Suppose $\varepsilon = 10$ and $2^{-\varepsilon} = 2^{-10} \approx 10^{-3}$, then m should be about 20 bits. So, if the size m of the target sub-key is small, then the approximation $q_\lambda \approx q$ may not be good. This puts into doubt the quality of normal approximation and hence of the applicability of the order statistics based method when m is small.

Constraint 2: $m - a$ must be large. Suppose that m is large enough so that $\lambda = M = 2^m - 1 \approx 2^m$ is large enough for the approximation $q_\lambda = q$ to be reasonable. We next consider the tightness of the upper bound given by Theorem 2.5. Under the approximation $q_\lambda \approx q$, denote the quantity Λ_{q_λ} given in (2.3) as Λ_q where

$$\Lambda_q = C \times \frac{q^2 + (1-q)^2}{(q(1-q))^{1/2}}.$$

Using $q = 1 - 2^{-a}$, the error bound $\Lambda_q/\lambda^{1/2}$ of Theorem 2.5 becomes

$$\frac{\Lambda_q}{\lambda^{1/2}} = C \times \frac{2^{-2a} + (1 - 2^{-a})^2}{(2^{-a}(1 - 2^{-a})(2^m - 1))^{1/2}} \approx C \times \frac{2^{-2a} + (1 - 2^{-a})^2}{(2^{m-a}(1 - 2^{-a}))^{1/2}}.$$

For the last quantity to be less than $2^{-\varepsilon}$ we must have

$$2^{m-a} \geq 2^{2\varepsilon} \times C^2 \times \frac{(2^{-2a} + (1 - 2^{-a})^2)^2}{1 - 2^{-a}}. \quad (4.1)$$

The maximum value of a is m and we may take the minimum value of a to be one, since attacks with advantage less than one are not really meaningful. For a in the range $[1, m]$, the value of $(2^{-2a} + (1 - 2^{-a})^2)/(1 - 2^{-a})$ is a small constant which gets close to 1 as a gets close to m . So, from (4.1) it follows that for the error in Theorem 2.5 to be less than $2^{-\varepsilon}$, it is required for $m - a$ to be around 2ε .

The value of a determines the advantage of the attack and a ranges from 1 to m . A 1-bit attack only eliminates half of the possible sub-key values, while an m -bit attack determines the correct key uniquely. So, the attack is more effective if the value of a is higher. However, from the above we see that for the normal approximation of the order statistics to be good, the value of a has to be relatively small compared to m . So, for attacks with high advantage, the analysis performed using this method is not applicable.

Remark 4.1. The quantity ε introduced in the above discussion is a bound on the quality of normal approximation, i.e., it is a bound on the error in approximating the actual distribution by a normal distribution. This quantity was not considered in earlier works on symmetric cipher cryptanalysis. We note in particular that ε does not denote the probability that the attack fails (i.e., the correct key is not recovered).

Remark 4.2. The value of $\varepsilon = 10$ indicates that the values of the two probabilities under consideration agree up to the second decimal place. This is a reasonable condition under which one value can be replaced by another. On the other hand, if one only wishes for agreement up to the first decimal place, then a value of $\varepsilon = 7$ will suffice. In this case, following the analysis done above, we will obtain that $m - a$ should be around 14. While 14 is lower than 20, this still indicates that the approximation is not meaningful for analysing attacks with high advantage.

Remark 4.3. It would be desirable to obtain an expression for the data complexity which involves ε . This, however, seems to be quite involved. Given the problems with the normal approximation of the order statistics, there does not seem to be too much motivation in following up on this approach. Further, the hypothesis testing framework allows for a much easier approach to the data complexity.

4.3 Selçuk's method of linear cryptanalysis: A critique

Let us briefly consider Selçuk's method of linear cryptanalysis [34]. Only one linear approximation was considered by Selçuk, i.e., $\ell = 1$. For simplicity, in this section, we will write L instead of L_1 and $L_{\kappa,j}$ instead of $L_{\kappa,j,1}$. Since there is a single linear approximation, the joint distribution $\tilde{p}$ reduces to simply a probability value $p = \Pr[L_{\kappa,j} = 0] \neq 1/2$ when κ is the correct choice. Selçuk assumes $p > 1/2$, the other case being similar. For an incorrect choice of κ , it is conventional to assume that $\Pr[L_{\kappa,j} = 0] = 1/2$.

Let $W_{\kappa} = (L_{\kappa,1} + \dots + L_{\kappa,N})/N - 1/2$. The test statistic corresponding to the choice κ of the sub-key is taken to be $T_{\kappa} = |W_{\kappa}|$. As mentioned earlier, for notational convenience, Selçuk considers T_0 to be the statistic corresponding to the correct key. So, the event corresponding to a successful attack with a -bit advantage is $T_0 > T_{(2^mq)}$ where $q = 1 - 2^{-a}$. The distribution of $T_{(2^mq)}$ is approximated to a normal distribution using order statistics as mentioned above.

Hidden assumption. Consider the distribution of T_0 . The random variable $L_{0,j}$ follows Bernoulli($1 - p$). Using Corollary 2.2, the distribution of W_0 can be approximated by the normal distribution $\mathcal{N}(p - 1/2, p(1 - p)/N)$. Since $T_0 = |W_0|$, it follows that T_0 approximately follows a folded normal distribution.

This, however, causes a problem in analysing the probability of the event $T_0 > T_{(2^mq)}$. With T_0 following a folded normal distribution and $T_{(2^mq)}$ approximately following a normal distribution, it becomes difficult to obtain the distribution of $T_0 - T_{(2^mq)}$.

Selçuk instead considers the event success to be $W_0 > T_{(2^mq)}$. No justification for this is provided. A careful consideration of this change of the success condition shows that there is a hidden assumption that the probability of $W_0 < 0$ is negligible. Indeed, if we do assume that $\Pr[W_0 < 0] \approx 0$, then we can compute as follows:

$$\begin{aligned} \Pr[T_0 > T_{(2^mq)}] &= \Pr[|W_0| > T_{(2^mq)}] \\ &= \Pr[W_0 > T_{(2^mq)} \mid W_0 > 0] \times \Pr[W_0 \geq 0] + \Pr[-W_0 > T_{(2^mq)} \mid W_0 < 0] \times \Pr[W_0 < 0] \\ &\approx \Pr[W_0 > T_{(2^mq)} \mid W_0 > 0] \times \Pr[W_0 \geq 0] \\ &= \Pr[W_0 > T_{(2^mq)}]. \end{aligned}$$

Without the assumption $\Pr[W_0 < 0] \approx 0$, we see no justification for considering success to be the event $W_0 > T_{(2^mq)}$. Having established that the condition $\Pr[W_0 < 0] \approx 0$ is required, we investigate the condition under which this approximation holds. Consider the event $W_0 < 0$. Recall that $c = 2(p - 1/2)$. Since we assume $p > 1/2$, it follows that $c > 0$. Then,

$$\Pr[W_0 < 0] = \Phi\left(-\frac{p - 1/2}{(p(1 - p)/N)^{1/2}}\right) = \Phi\left(-\frac{c\sqrt{N}}{\sqrt{1 - c^2}}\right).$$

So, N has to be sufficiently large to ensure that the probability of $W_0 < 0$ is negligible. For example, $\Phi(-5) \approx 2.86651571879 \times 10^{-7}$. If we consider the last value to be small enough, it is required to choose N such that $c\sqrt{N}/\sqrt{1 - c^2} > 5$ which translates to the condition $N > 25(1 - c^2)/c^2$. This bound is to be compared with the expression for N (see (6.5)) obtained by Selçuk [34] from analysing the condition for success.

Implicit reference to the goodness of approximation. Selçuk mentions an assumption. The random variable $T_{(2^m q)}$ is non-negative by definition. Using the order statistics result, its distribution is approximated by a normal $\mathcal{N}(\mu_q, \sigma_q^2)$ distribution, where the mean μ_q and variance σ_q^2 depend on q . The normal distribution, however, can take negative values. Selçuk puts the condition that $\Pr[T_{(2^m q)} < 0] = \Phi(-\mu_q/\sigma_q) \approx 0$. The value $\Phi(-5)$ is considered to be negligible giving rise to the condition $\mu_q/\sigma_q > 5$. It is mentioned in [34], that this holds if a and m are both greater than 8. This puts a lower bound on the size of the key that can be attacked.

The above condition of requiring $\Pr[T_{(2^m q)} < 0] \approx 0$ is somewhat artificial. It is certainly not required for analysing the success probability of the attack. The appropriate issue is that the approximation of the distribution of $T_{(2^m q)}$ by a normal distribution should be a good one. By definition, $T_{(2^m q)}$ is non-negative and so if the approximation is good, then $\Pr[T_{(2^m q)} < 0]$ will be negligible. On the other hand, it is possible for $\Pr[T_{(2^m q)} < 0]$ to be negligible and yet the approximation of the distribution of $T_{(2^m q)}$ by the normal approximation may not be good. This point is missing in [34].

5 Key recovery attack via hypothesis testing

For each choice κ of the target sub-key, the test statistic T_κ is defined. The distribution of T_κ depends on whether κ is the correct key choice or whether it is an incorrect key choice. This leads to two distributions: $\mathcal{D}_0$ corresponding to the correct key choice and $\mathcal{D}_1$ corresponding to an incorrect key choice.

The attack can be formulated as a hypothesis testing framework. Given a choice of κ , the statistic T_κ is used to determine whether κ is a possible candidate key or not. The null hypothesis H_0 is tested against the alternate hypothesis H_1 . The structure of the test and the decision rule are as follows:

$$\begin{cases} H_0: \kappa \text{ is correct} & \text{versus} & H_1: \kappa \text{ is incorrect.} \\ \text{Decision rule: Reject } H_0 & \text{if } T_\kappa \leq t. \end{cases} \quad (5.1)$$

Here t is a threshold parameter to be determined later. The hypothesis testing examples that we consider later will primarily be of the above form. (In some cases, the test can also take the form $T_\kappa \geq t$ and the analysis in this case is similar.)

As usual, a successful attack corresponds to the event that κ is accepted when H_0 holds. We denote this event by succ . We define three quantities α , β and P_S based on Type-I and Type-II errors as follows:

$$\begin{aligned} \Pr[\text{Type-I error}] &= \Pr[\kappa \text{ is rejected} \mid H_0 \text{ holds}] = \Pr[T \leq t \mid H_0 \text{ holds}] \leq \alpha, \\ \Pr[\text{Type-II error}] &= \Pr[\kappa \text{ is accepted} \mid H_1 \text{ holds}] = \Pr[T > t \mid H_1 \text{ holds}] \leq \beta, \\ \Pr[\text{succ}] &= 1 - \Pr[\text{Type-I error}] \geq 1 - \alpha = P_S. \end{aligned}$$

In some cases, we will have α and β to be respectively equal to the probabilities of Type-I and Type-II errors. In certain other cases, though, it is more convenient to have α and β to be appropriate upper bounds for the probabilities of Type-I and Type-II errors. If α is an upper bound on the probability of Type-I error, then P_S is a lower bound on the actual probability of success.

The analysis then proceeds in the following manner. From the distributions $\mathcal{D}_0$ and $\mathcal{D}_1$ it is required to compute expressions for α (and hence of $P_S = 1 - \alpha$) and β . These expressions involve the threshold parameter t and the data complexity N . Eliminating N between these two expressions provides the expression for the threshold parameter t in terms of P_S and β . On the other hand, eliminating t between these two expressions provides an expression for the data complexity N in terms of P_S and β .

Relating Type-II error probability to advantage. It is possible to relate the data complexity N to the expected advantage of the attack. Whenever a Type-II error occurs, an incorrect key gets listed as a possible candidate key. This event occurs with probability β . The total number of choices for κ is 2^m out of which one is correct and the others are incorrect. So, the expected number of incorrect keys which gets listed as candidate keys is $\beta(2^m - 1) \approx \beta 2^m$. For an a -bit advantage attack, the size of the false alarm set is 2^{m-a} . Setting $\beta 2^m = 2^{m-a}$, we obtain

$$\beta = 2^{-a}. \quad (5.2)$$

Substituting $\beta = 2^{-a}$ in the expression for N provides the data complexity in terms of the success probability P_S and the (expected) advantage a .

In the following sections we illustrate this methodology by applying it on a number of cases of known linear and differential attacks.

6 Linear cryptanalysis

As mentioned earlier, the order statistics based approach has been used to analyse data complexity of key recovery attacks. The case of single linear cryptanalysis was considered in [34] and that of multiple linear cryptanalysis was considered in [20]. The analysis of Section 4 puts a question mark on the framework of the order statistics based approach. In this section, we show that the data complexity results can be derived using the hypothesis testing framework and thus avoid the several difficulties with the order statistics based framework.

6.1 Single linear cryptanalysis

In this section, we consider the case where a single linear approximation is available. The data complexity of a key recovery attack using such an approximation was derived in [34] using the order statistics based approach. Here, we derive the same data complexity using the hypothesis testing framework and so avoid the various pitfalls that are inherent in the approximations required for the order statistics based approach. The setting and notation for this section are as described in Section 4.3.

For each $j \in \{1, \dots, N\}$, under the null hypothesis H_0 (i.e., for the correct choice of κ), $L_{\kappa,j}$ follows $\text{Bernoulli}(p)$ for each j . Here $p = 1/2 \pm |\epsilon|$ where the sign is determined by the inner key bit. The actual value of the sign will not matter in the ensuing analysis. Under the alternate hypothesis H_1 (i.e., for an incorrect choice of κ), $L_{\kappa,j}$ follows $\text{Bernoulli}(1/2)$.

Let $\mu_0 = p$, $\sigma_0^2 = p(1-p)$ and $\mu_1 = 1/2$, $\sigma_1^2 = 1/4$ denote the mean and the variance of $L_{\kappa,j}$ under H_0 and H_1 , respectively. Let $c = 2(p - 1/2)$ so that $|\mu_0 - \mu_1| = |c|$ and $\sigma_0^2 = 1 - c^2$.

Define the test statistic T_κ in the following manner:

$$T_\kappa = |Z_\kappa|, \quad Z_\kappa = \frac{(L_{\kappa,1} + \dots + L_{\kappa,N})/N - \mu_1}{\sigma_1/\sqrt{N}}.$$

The hypothesis testing framework and the decision rule mentioned in (5.1) are applied to T_κ .

From Corollary 2.2, under H_1 , the distribution for Z_κ is approximated by the standard normal distribution. Since $\mu_1 = 1/2$, the error bound is $C/N^{1/2}$. If we require the error in approximation to be at most $2^{-\epsilon}$, then it is sufficient to have $N > 2^{2\epsilon}C^2$.

We write

$$Z_\kappa = \frac{\sigma_0}{\sigma_1} W_\kappa + \sqrt{N} \left(\frac{\mu_0 - \mu_1}{\sigma_1} \right), \quad W_\kappa = \frac{(L_{\kappa,1} + \dots + L_{\kappa,N})/N - \mu_0}{\sigma_0/\sqrt{N}}.$$

Again from Corollary 2.2, under H_0 , the distribution of W_κ is approximated by the standard normal distribution. In this case, since $\mu_0 = p \neq 1/2$, the error bound is $C(1 + c^2)/((1 - c^2)N)^{1/2}$. For

$$N \geq 2^{2\epsilon}C^2 \times \frac{(1 + c^2)^2}{1 - c^2} \quad (6.1)$$

the error in approximation will be at most $2^{-\epsilon}$. If p is close to $1/2$ (which is often the case) and so c is close to 0, then $N \approx 2^{2\epsilon}$ ensures that the approximation error is around $2^{-\epsilon}$. More generally, one has to consider the bound given by (6.1).

The Type-I and Type-II error probabilities can be computed as follows:

$$\begin{aligned} \Pr[\text{Type-I error}] &= \Pr[T_\kappa \leq t \mid H_0 \text{ holds}] \\ &= \Pr \left[-\left(\frac{\sigma_1 t}{\sigma_0} + \sqrt{N} \left(\frac{\mu_0 - \mu_1}{\sigma_0} \right) \right) \leq W_\kappa \leq \frac{\sigma_1 t}{\sigma_0} - \sqrt{N} \left(\frac{\mu_0 - \mu_1}{\sigma_0} \right) \mid H_0 \text{ holds} \right]. \end{aligned}$$

Suppose that $\mu_0 > \mu_1$. In this case

$$\begin{aligned} \Pr[\text{Type-I error}] &= \Pr\left[-\left(\frac{\sigma_1 t}{\sigma_0} + \frac{|\mu_0 - \mu_1|\sqrt{N}}{\sigma_0}\right) \leq W_\kappa \leq \frac{\sigma_1 t}{\sigma_0} - \frac{|\mu_0 - \mu_1|\sqrt{N}}{\sigma_0} \mid H_0 \text{ holds}\right] \\ &\leq \Pr\left[-\infty \leq W_\kappa \leq \frac{\sigma_1 t}{\sigma_0} - \frac{|\mu_0 - \mu_1|\sqrt{N}}{\sigma_0} \mid H_0 \text{ holds}\right] = \Phi^{-1}\left(\frac{\sigma_1 t}{\sigma_0} - \frac{|\mu_0 - \mu_1|\sqrt{N}}{\sigma_0}\right). \end{aligned}$$

On the other hand, if $\mu_0 < \mu_1$, then

$$\begin{aligned} \Pr[\text{Type-I error}] &= \Pr\left[-\left(\frac{\sigma_1 t}{\sigma_0} - \frac{|\mu_0 - \mu_1|\sqrt{N}}{\sigma_0}\right) \leq W_\kappa \leq \frac{\sigma_1 t}{\sigma_0} + \frac{|\mu_0 - \mu_1|\sqrt{N}}{\sigma_0} \mid H_0 \text{ holds}\right] \\ &\leq \Pr\left[-\left(\frac{\sigma_1 t}{\sigma_0} - \frac{|\mu_0 - \mu_1|\sqrt{N}}{\sigma_0}\right) \leq W_\kappa \leq \infty \mid H_0 \text{ holds}\right] \\ &= 1 - \Phi\left(-\left(\frac{\sigma_1 t}{\sigma_0} - \frac{|\mu_0 - \mu_1|\sqrt{N}}{\sigma_0}\right)\right) = \Phi\left(\frac{\sigma_1 t}{\sigma_0} - \frac{|\mu_0 - \mu_1|\sqrt{N}}{\sigma_0}\right). \end{aligned}$$

We set

$$\alpha = \Phi\left(\frac{\sigma_1 t}{\sigma_0} - \frac{|\mu_0 - \mu_1|\sqrt{N}}{\sigma_0}\right) \quad \text{and} \quad P_S = 1 - \alpha. \quad (6.2)$$

So, for any $\mu_0 \neq \mu_1$, as $\Pr[\text{Type-I error}] \leq \alpha$ and $P_S = 1 - \alpha$, the probability of success is at least P_S . Using (6.2), we have

$$\sigma_1 t = \sigma_0 \Phi^{-1}(1 - P_S) + |\mu_0 - \mu_1|\sqrt{N} = -\sigma_0 \Phi^{-1}(P_S) + |\mu_0 - \mu_1|\sqrt{N}. \quad (6.3)$$

This allows the threshold parameter t to be expressed in terms of P_S . Further, we set

$$\begin{aligned} \beta &= \Pr[\text{Type-II error}] = \Pr[|T_\kappa| > t \mid H_1 \text{ holds}] \\ &= \Pr[T_\kappa < -t \mid H_1 \text{ holds}] + \Pr[T_\kappa > t \mid H_1 \text{ holds}] \\ &= \Phi(-t) + 1 - \Phi(t) = 2(1 - \Phi(t)). \end{aligned} \quad (6.4)$$

Eliminating t using (6.3) and (6.4), substituting the values of σ_0 , σ_1 and $(\mu_0 - \mu_1)^2$ in terms of c , and using (5.2), we obtain

$$N = \frac{\{\sigma_1 \Phi^{-1}(1 - \beta/2) + \sigma_0 \Phi^{-1}(P_S)\}^2}{(\mu_0 - \mu_1)^2} = \frac{\{\Phi^{-1}(1 - 2^{-a-1}) + \sqrt{1 - c^2} \Phi^{-1}(P_S)\}^2}{c^2}. \quad (6.5)$$

In [34], Selçuk makes the assumption $1 - c^2 \approx 1$. Using this assumption, the expression for N given by (6.5) becomes exactly the same as that obtained in [34].

It will usually be the case that for a reasonable value of ε , the expression for N given by (6.5) is larger than the bound given by (6.1). This will ensure that the approximation is good enough for the analysis to go through. For a concrete situation, one has to plug in the values of ε , p , a and P_S and then compare the expressions for N given by (6.5) and (6.1).

6.2 Multiple linear cryptanalysis

In this section, $\ell \geq 1$ and we assume the setting and the notation explained in Section 3.1. Recall that for a choice κ of the target sub-key and $\eta \in \{0, 1\}^\ell$, the random variable $Q_{\kappa, \eta}$ has been defined in (3.4) to be $Q_{\kappa, \eta} = \#\{j \in \{1, \dots, N\} : X_{\kappa, j} = \eta\}$, i.e., $Q_{\kappa, \eta}$ is the number of times η appears among the random variables $X_{\kappa, 1}, \dots, X_{\kappa, N}$.

6.2.1 Data complexity for the χ^2 test statistics

The test statistics used by Hermelin, Cho and Nyberg [20] was

$$T_\kappa = 2^\ell N \sum_{\eta=0}^{2^\ell-1} (Q_{\kappa, \eta}/N - 2^{-\ell})^2. \quad (6.6)$$

We use T_κ in the hypothesis testing framework given in (5.1). The distribution of T_κ is approximated in [20] as follows:

$$T_\kappa \sim \begin{cases} \chi_{2^\ell-1}^2(NC(\tilde{p})) & \text{under } H_0, \\ \chi_{2^\ell-1}^2 & \text{under } H_1. \end{cases} \quad (6.7)$$

This is justified by [15]; later we consider this issue in more details. Since it is difficult to work with a non-central chi-squared distribution, Hermelin, Cho and Nyberg [20] approximate $\chi_{2^\ell-1}^2(NC(\tilde{p}))$ for large values of $2^\ell - 1 + NC(\tilde{p})$ by normal $\mathcal{N}(2^\ell - 1 + NC(\tilde{p}), 2(2^\ell - 1 + 2NC(\tilde{p})))$ distribution. We will also take a close look at this approximation later.

Let $\Psi_{2^\ell-1}(x)$ be the distribution function of the central $\chi_{2^\ell-1}^2$ distribution. Then

$$\begin{cases} \alpha = \Pr[\text{Type-I error}] = \Pr[T_\kappa \leq t \mid H_0 \text{ holds}] = \Pr[T_\kappa \leq t] = \Phi\left(\frac{t - (2^\ell - 1 + NC(\tilde{p}))}{\sqrt{2(2^\ell - 1 + 2NC(\tilde{p}))}}\right) = 1 - P_S, \\ \beta = \Pr[\text{Type-II error}] = \Pr[T_\kappa > t \mid H_1 \text{ holds}] = 1 - \Psi_{2^\ell-1}(t). \end{cases} \quad (6.8)$$

So,

$$t = 2^\ell - 1 + NC(\tilde{p}) + \sqrt{2(2^\ell - 1 + 2NC(\tilde{p}))}\Phi^{-1}(1 - P_S) = \Psi_{2^\ell-1}^{-1}(1 - \beta).$$

Simplifying leads to the following relation which is a quadratic equation in $(NC(\tilde{p}))$:

$$(NC(\tilde{p}))^2 - 2(NC(\tilde{p}))\{\gamma - 2^\ell - 1 + 2\delta^2\} + \{\gamma - 2^\ell - 1\}^2 - 2(2^\ell - 1)\delta^2 = 0,$$

where $\gamma = \Psi_{2^\ell-1}^{-1}(1 - \beta)$ and $\delta = \Phi^{-1}(1 - P_S) = -\Phi^{-1}(P_S)$. A routine calculation shows that the discriminant of this quadratic is positive and so there are real solutions for $NC(\tilde{p})$. Taking the positive square root of the discriminant and using (5.2) to substitute 2^{-a} for β , we obtain

$$N = \frac{1}{C(\tilde{p})} \left[2\{\Phi^{-1}(P_S)\}^2 + \Psi_{2^\ell-1}^{-1}(1 - 2^{-a}) - (2^\ell - 1) + |\Phi^{-1}(P_S)|\sqrt{2\Psi_{2^\ell-1}^{-1}(1 - 2^{-a}) - 4\{\Phi^{-1}(P_S)\}^2} \right]. \quad (6.9)$$

Approximating $\chi_{2^\ell-1}^2$ by $\mathcal{N}(2^\ell - 1, 2(2^\ell - 1))$. The expression for N given by (6.9) involves both Φ and Ψ . This is not a problem, since both these distribution functions can be simulated. On the other hand, let us consider the effect of approximating $\chi_{2^\ell-1}^2$ by $\mathcal{N}(2^\ell - 1, 2(2^\ell - 1))$ as has been done in [20]. The expression for β now becomes

$$\beta = \Pr[\text{Type-II error}] = 1 - \Phi\left(\frac{t - (2^\ell - 1)}{\sqrt{2(2^\ell - 1)}}\right).$$

Combining this with the expression for P_S given by (6.8), we obtain the following quadratic in $(NC(\tilde{p}))$:

$$(NC(\tilde{p}))^2 - 2(NC(\tilde{p}))\left(\sqrt{2(2^\ell - 1)}\gamma + 2\delta^2\right) + 2(2^\ell - 1)(\gamma^2 - \delta^2) = 0, \quad (6.10)$$

where now $\gamma = \Phi^{-1}(1 - \beta)$ and δ is equal to $\Phi^{-1}(1 - P_S) = -\Phi^{-1}(P_S)$ as before. Again, a routine calculation shows that the discriminant of the quadratic is positive showing that there are real solutions for $NC(\tilde{p})$. Taking the positive root of the discriminant and using (5.2) to substitute 2^{-a} for β , we obtain

$$\begin{aligned} N = \frac{1}{C(\tilde{p})} & \left[\sqrt{2(2^\ell - 1)}\Phi^{-1}(1 - 2^{-a}) + 2(\Phi^{-1}(P_S))^2 \right. \\ & \left. + \Phi^{-1}(P_S)\sqrt{4(\Phi^{-1}(P_S))^2 + 4\sqrt{2(2^\ell - 1)}\Phi^{-1}(1 - 2^{-a}) + 2(2^\ell - 1)} \right]. \end{aligned} \quad (6.11)$$

As mentioned in [20], this approximation is valid only for $\ell > 5$. Later we provide a much more detailed analysis of the approximation.

Comparison to the data complexity obtained in [20]. Equation (9) of [20] gives the success probability of Selçuk's method to be

$$P_S = \Phi\left(\frac{\mu_R - \mu_q}{(\sigma_R^2 + \sigma_q^2)^{1/2}}\right).$$

From [20, Section 5.1],

$$\mu_R = (2^\ell - 1) + NC(\tilde{p}), \quad \sigma_R^2 = 2((2^\ell - 1) + 2NC(\tilde{p})), \quad \mu_q = \sqrt{2(2^\ell - 1)\Phi^{-1}(1 - 2^{-a}) + (2^\ell - 1)},$$

and

$$\sigma_q^2 = \frac{2^{-(\ell+a)}\sigma_w^2}{\phi(\Phi^{-1}(1 - 2^{-a}))} \ll \sigma_R^2.$$

Using these values in the expression for P_S and simplifying leads to a quadratic in $NC(\tilde{p})$ which turns out to be exactly the same as the one in (6.10). Consequently, the expression for the data complexity N in this case is also given by (6.11). In [20], Hermelin, Cho and Nyberg applied some further approximations to (6.11) to arrive at a simplified expression for the data complexity. They noted that such approximations are valid for large P_S and large advantage a . We did not investigate these approximations any further.

6.2.2 Data complexity for the LLR test statistics

Recall that the joint distribution of the L_i is parameterised by z which is the tuple representing the inner key bits. So, for every $\eta \in \{0, 1\}^\ell$, $\tilde{p}_z(\eta)$ is the probability that $(L_1 \oplus z_1, \dots, L_\ell \oplus z_\ell) = \eta$. For a choice κ of the last round key bits and z of the inner key bits we make the following definitions:

$$X_{\kappa,z,j} = (L_{\kappa,j,1} \oplus z_1, \dots, L_{\kappa,j,\ell} \oplus z_\ell), \\ Q_{\kappa,z,\eta} = \#\{j : X_{\kappa,z,j} = \eta\} \quad \text{for } \eta = 0, \dots, 2^\ell - 1.$$

In this case, the test statistics is determined by both the choice of the last round key bits κ and the choice of the inner key bits z .

The hypothesis testing framework for this case varies a little from what has been described in Section 5. The null hypothesis H_0 and the alternate hypothesis H_1 remain unchanged. The test statistic and the decision rule change. We denote the test statistics by $T_{\kappa,z}$ which is defined as follows:

$$T_{\kappa,z} = \sum_{\eta \in \{0,1\}^\ell} Q_{\kappa,z,\eta} \ln\left(\frac{\tilde{p}_z(\eta)}{2^{-\ell}}\right). \quad (6.12)$$

This is exactly the LLR test statistics considered in [20]. The test statistic $T_{\kappa,z}$ approximately follows $\mathcal{N}(N\mu, N\sigma^2)$, where

$$\mu = \begin{cases} \mu_0 = D(\tilde{p}_z \parallel \tilde{p}_\$) & \text{if } H_0 \text{ holds,} \\ \mu_1 = -D(\tilde{p}_\$ \parallel \tilde{p}_z) & \text{if } H_1 \text{ holds,} \end{cases} \quad (6.13)$$

$$\sigma^2 = \begin{cases} \sigma_0^2 = \sum_{\eta=1}^v p_z(\eta) (\ln(\frac{p_z(\eta)}{p_\$(\eta)}))^2 - \mu_0^2 & \text{if } H_0 \text{ holds,} \\ \sigma_1^2 = \sum_{\eta=1}^v p_\$(\eta) (\ln(\frac{p_\$(\eta)}{p_z(\eta)}))^2 - \mu_1^2 & \text{if } H_1 \text{ holds.} \end{cases} \quad (6.14)$$

Here

$$D(\tilde{p} \parallel \tilde{q}) = \sum_{\eta} \tilde{p}(\eta) \ln\left(\frac{\tilde{p}(\eta)}{\tilde{q}(\eta)}\right)$$

denotes the Kullback–Leibler divergence between the two probability distributions $\tilde{p}$ and $\tilde{q}$. From (6.13) and (6.14), it appears that μ and σ depend on z . We later show that this is not the case and that the values of μ and σ are the same for all z . Also, later we consider the error in the normal approximation.

Note that μ_0 is positive whereas μ_1 is negative. For a threshold parameter t the exact value of which is to be determined. We formulate the following decision rule:

$$\text{Reject } H_0 \text{ if } T_{\kappa,z} \leq t \text{ for all } z \in \{0, 1\}^\ell.$$

Since $\mu_1 < 0 < \mu_0$, for the test to be meaningful, the value of t should be in $(N\mu_1, N\mu_0)$.

Suppose z^* is the correct choice of the inner key bits. The probability of Type-I error is

$$\Pr[T_{\kappa,z} \leq t \text{ for all } z \in \{0, 1\}^\ell \mid H_0 \text{ holds}] \leq \Pr[T_{\kappa,z^*} \leq t \mid H_0 \text{ holds}].$$

We define $\alpha = \Pr[T_{\kappa,z^*} \leq t \mid H_0 \text{ holds}]$ and so α is an upper bound on the probability of Type-I error. Also, we define $P_S = 1 - \alpha$ so that P_S is a lower bound on the success probability.

Since z^* is the correct choice of the inner key bits, the distribution p_{z^*} is the correct distribution and so under H_0 , T_{κ,z^*} approximately follows $\mathcal{N}(N\mu_0, N\sigma_0^2)$. From this we obtain

$$1 - P_S = \alpha = \Pr[T_{\kappa,z^*} \leq t \mid H_0 \text{ holds}] = \Phi\left(\frac{t - N\mu_0}{\sqrt{N}\sigma_0}\right) \Rightarrow t = N\mu_0 + \sqrt{N}\sigma_0\Phi^{-1}(1 - P_S). \quad (6.15)$$

Remark 6.1. Note that if $P_S \leq 0.5$, then $t \geq N\mu_0$ which violates the condition that t should be in $(N\mu_0, N\mu_1)$. So, the condition $P_S > 0.5$ is required for the test to be meaningful. This is an implicit consequence of the approximation of the test statistic by a normal distribution. A similar condition on the success probability has been mentioned in [20, Theorem 2] though no explanation was provided.

Under H_1 , $T_{\kappa,z}$ approximately follows $\mathcal{N}(N\mu_1, N\sigma_1^2)$ irrespective of the choice of z . The probability of Type-II error is

$$\begin{aligned} \Pr[T_{\kappa,z} > t \text{ for some } z \mid H_1 \text{ holds}] &\leq \sum_{z \in \{0,1\}^\ell} \Pr[T_{\kappa,z} > t \mid H_1 \text{ holds}] \\ &= 2^\ell (1 - \Pr[T_{\kappa,z} \leq t \mid H_1 \text{ holds}]) \\ &= 2^\ell \left(1 - \Phi\left(\frac{t - N\mu_1}{\sqrt{N}\sigma_1}\right)\right). \end{aligned}$$

We set

$$\beta = 2^\ell \left(1 - \Phi\left(\frac{t - N\mu_1}{\sqrt{N}\sigma_1}\right)\right)$$

so that β is an upper bound on the probability of Type-II error. Eliminating t from the expression for β and (6.15) and using (5.2) to substitute 2^{-a} for β provides the following expression for the data complexity N :

$$N = \frac{\{\sigma_1\Phi^{-1}(1 - 2^{-\ell-a}) + \sigma_0\Phi^{-1}(P_S)\}^2}{(\mu_0 - \mu_1)^2}. \quad (6.16)$$

Baignères, Junod and Vaudenay [2] mentioned that the values of μ_0 , μ_1 , σ_0 and σ_1 can be approximated as follows:

$$\mu = \begin{cases} \mu_0 \approx \frac{1}{2}C(\tilde{p}) & \text{if } H_0 \text{ holds,} \\ \mu_1 \approx -\frac{1}{2}C(\tilde{p}) & \text{if } H_1 \text{ holds,} \end{cases} \quad \text{and} \quad \sigma_0^2, \sigma_1^2 \approx C(\tilde{p}).$$

This approximation causes certain problems which we discuss later. Using these values in (6.16), we obtain the following expression for the data complexity:

$$N = \frac{\{\Phi^{-1}(1 - 2^{-\ell-a}) + \Phi^{-1}(P_S)\}^2}{C(\tilde{p})}. \quad (6.17)$$

Comparison to the data complexity obtained in [20]. By [20, (38)] the data complexity of the LLR method is given by

$$\frac{\{\Phi^{-1}(P_{12}) + \Phi^{-1}(\sqrt[2^\ell]{1 - 2^{-a}})\}^2}{C(\tilde{p})}. \quad (6.18)$$

Here P_{12} is a lower bound on the success probability and can be taken to be P_S . In [20], Hermelin, Cho and Nyberg used the approximation $\sqrt[2^\ell]{1 - 2^{-a}} \approx 1 - 2^{-a-\ell}$. Using this approximation in (6.18) gives the expression for N to be exactly the same as that given in (6.17).

7 Differential cryptanalysis

Expressions for the data complexities of differential cryptanalysis were derived in [34] and [11] using the order statistics based approach. In this section, we show that these expressions can be derived using the hypothesis testing based framework.

7.1 Single differential cryptanalysis

We present a brief description of the setting of differential cryptanalysis and follow it up by the analysis of the data complexity of single differential cryptanalysis.

Let $\delta_0, \delta_1, \dots, \delta_r$ be n -bit strings where δ_0 is not the all-zero string. For a uniform random P , set $P' = P \oplus \delta_0$. Let $B^{(0)} = P, B^{(1)}, \dots, B^{(r)}$ be the inputs to round numbers $0, \dots, r$ respectively, i.e., $B^{(i+1)} = R_{k^{(i)}}^{(i)}(B^{(i)})$ and suppose that $B^{(0)'} = P', B^{(1)'}, \dots, B^{(r)'}$ are the inputs to round numbers $0, \dots, r$, respectively, corresponding to the plaintext P' . Consider the event $A = \bigwedge_{i=0}^r (B^{(i)} \oplus B^{(i)'} = \delta_i)$ and suppose that for the correct key K , $\Pr[A] = p$. Here the randomness and the probability is over the uniform random choice of P .

As in Section 3, we assume that guessing m bits of the key allows the partial decryption of C to obtain $B^{(r)}$. These m bits will constitute the target sub-key and the goal will be to obtain the correct value of the sub-key. We will denote a choice of the target sub-key by κ .

The initial study of the block cipher uncovers the masks $\delta_0, \delta_1, \dots, \delta_r$, the probability p and the target sub-key bits. As in the case of linear cryptanalysis, this is a non-trivial task and forms the foundation for a statistical analysis of the attack.

Suppose that we have obtained the ciphertexts C and C' corresponding to the plaintexts P and P' , respectively. Guessing a target sub-key allows computing $B^{(r)}$ and $B^{(r)'}$ from C and C' , respectively. The earlier $B^{(i)}$ and $B^{(i)'}$ cannot be obtained. Define the event D to be $B^{(r)} \oplus B^{(r)'} = \delta_r$. The event A implies the event D . The event D may also occur even when $\bar{A}$ occurs. In other words, A defines one possible path for the event D to occur while there may be other paths which also lead to D . Let $\Pr[D | \bar{A}] = p'$ and let $p_0 = p + (1 - p)p'$. Then for the correct choice κ of the target sub-key, we have

$$\begin{aligned} \Pr[D] &= \Pr[D \wedge (A \vee \bar{A})] \\ &= \Pr[D | A] \Pr[A] + \Pr[D | \bar{A}] \Pr[\bar{A}] \\ &= p + p'(1 - p) = p_0. \end{aligned}$$

Since δ_0 is not the zero string, we have $P \neq P'$. Since each round function is a bijection, we have $B^{(i)} \neq B^{(i)'}$ for $i = 1, \dots, r$. If the choice κ of the target sub-key is incorrect, then it is reasonable to assume that $B^{(r)}$ and $B^{(r)'}$ correspond to uniform sampling without replacement of two n -bit strings from $\{0, 1\}^n$. So, if κ is an incorrect choice of the target sub-key, then $\Pr[D] = 1/(2^n - 1)$. Let $p_w = 1/(2^n - 1)$. In general $p_0 > p_w$.

For the actual attack, there will be N matched pairs of plaintexts $(P_1, P'_1), \dots, (P_N, P'_N)$ with $P_j \oplus P'_j = \delta_0$ and their corresponding ciphertexts $(C_1, C'_1), \dots, (C_N, C'_N)$. The randomness arises for the independent and uniform distribution of $P_1, \dots, P_N$. For a choice κ of the target sub-key, we obtain

$$(B_{\kappa,1}^{(r)}, B_{\kappa,1}^{(r)'}) , \dots , (B_{\kappa,N}^{(r)}, B_{\kappa,N}^{(r)'})$$

by inverting $(C_1, C'_1), \dots, (C_N, C'_N)$, respectively. So, it is possible to determine whether the condition $B_{\kappa,j}^{(r)} \oplus B_{\kappa,j}^{(r)'} = \delta_r$ holds for each $j = 1, \dots, N$.

For a choice κ of the target sub-key and $j = 1, \dots, N$, define random variables

$$W_{\kappa,j} = \begin{cases} 1 & \text{if } B_{\kappa,j}^{(r)} \oplus B_{\kappa,j}^{(r)'} = \delta_r, \\ 0 & \text{otherwise.} \end{cases}$$

If κ is the correct choice, then $\Pr[W_{\kappa,j} = 1] = p_0$. If κ is an incorrect choice, then $\Pr[W_{\kappa,j} = 1] = p_w$ for all j .

The hypothesis testing framework in (5.1) is to be applied. Define the test statistics

$$T_k = \frac{(W_{k,1} + \cdots + W_{k,N}) - Np_w}{\sqrt{Np_w(1-p_w)}}.$$

From Corollary 2.2, under H_1 , T_k is approximated by the standard normal distribution for sufficiently large N . From the error bound, we have that if

$$N \geq 2^{2\varepsilon} C^2 \frac{(p_w^2 + (1-p_w)^2)^2}{p_w(1-p_w)}, \quad (7.1)$$

then the error in approximation is at most $2^{-\varepsilon}$. Write

$$T_k = \frac{(W_{k,1} + \cdots + W_{k,N}) - Np_w}{\sqrt{Np_w(1-p_w)}} = Y_k \times \frac{\sqrt{Np_0(1-p_0)}}{\sqrt{Np_w(1-p_w)}} + \frac{Np_0 - Np_w}{\sqrt{Np_w(1-p_w)}},$$

where

$$Y_k = \frac{(W_{k,1} + \cdots + W_{k,N}) - Np_0}{\sqrt{Np_0(1-p_0)}}.$$

Again from Corollary 2.2, under H_0 , the distribution of Y_k is approximated by the standard normal distribution. Further, if

$$N \geq 2^{2\varepsilon} C^2 \frac{(p_0^2 + (1-p_0)^2)^2}{p_0(1-p_0)}, \quad (7.2)$$

then the error in approximation is at most $2^{-\varepsilon}$. Using this approximation, we have

$$\begin{aligned} \alpha &= \Pr[\text{Type-I Error}] = \Pr[T_k \leq t \mid H_0 \text{ holds}] \\ &= \Pr\left[Y_k \leq \left(\frac{\sqrt{Np_w(1-p_w)}}{\sqrt{Np_0(1-p_0)}}\right)\left(t - \frac{Np_0 - Np_w}{\sqrt{Np_w(1-p_w)}}\right)\right] \\ &= \Pr\left[Y_k \leq \frac{t\sqrt{Np_w(1-p_w)}}{\sqrt{Np_0(1-p_0)}} - \frac{Np_0 - Np_w}{\sqrt{Np_0(1-p_0)}}\right] \\ &= \Phi\left(\frac{t\sqrt{Np_w(1-p_w)}}{\sqrt{Np_0(1-p_0)}} - \frac{Np_0 - Np_w}{\sqrt{Np_0(1-p_0)}}\right) = 1 - P_S. \end{aligned} \quad (7.3)$$

Type-II error is given by

$$\beta = \Pr[\text{Type-II Error}] = \Pr[T_k > t \mid H_1 \text{ holds}] = 1 - \Phi(t),$$

which implies $t = \Phi^{-1}(1 - \beta)$. Using this value of t in (7.3) and using (5.2) to substitute 2^{-a} for β , we obtain

$$N = \left\{ \frac{\sqrt{p_w(1-p_w)}\Phi^{-1}(1-2^{-a}) + \sqrt{p_0(1-p_0)}\Phi^{-1}(P_S)}{p_0 - p_w} \right\}^2.$$

Selçuk [34] used the assumptions that $p_w = p_r$, $1 - p \approx 1$, $1 - p_0 \approx 1$ and $1 - p_w \approx 1$. Using these relations, we get

$$N = \left\{ \frac{\sqrt{p_w}\Phi^{-1}(1-2^{-a}) + \sqrt{p_0}\Phi^{-1}(P_S)}{p_0 - p_w} \right\}^2. \quad (7.4)$$

This is identical to the expression for data complexity obtained by Selçuk in [34] using the order statistics based approach.

The expression for N given by (7.4) has to be compared with the lower bounds for N given by (7.1) and (7.2). Usually, it will be the case that $p_0 > p_w$ and so it is sufficient to only consider the lower bound given by (7.1). The right-hand side of (7.4) is inversely proportional to $(p_0 - p_w)^2$ whereas the right-hand side of (7.1) is inversely proportional to p_w . Since in practice it will usually be the case that $|p_0 - p_w|$ is less than p_w , one would expect the expression given by (7.4) to be greater than the one given by (7.1). The actual comparison of the two expressions can be obtained only by plugging in the relevant values of ε , p_w , p_0 , a and P_S .

Selçuk [34] remarked that the standard normal approximations required in this context need not be good. The concrete error analysis given here was not done. In fact, from our analysis, it appears that in many usual cases the expression for the final data complexity given by (7.4) will be large enough for the normal approximations to be proper.

7.2 Multiple differential cryptanalysis

In the case of multiple differential attacks, it is assumed that the prior analysis of the block cipher results in v sequences of n -bit strings $\delta_j^{(1)}, \dots, \delta_j^{(v)}$ for $j = 1, \dots, r$. Assume as before that there is an m -bit target sub-key and the goal of the attack is to obtain the correct value for this sub-key. Further, the quantities (P_j, P'_j) , (C_j, C'_j) , $j = 1, \dots, N$ are defined as in the case of a single differential attack, i.e., $P_j \oplus P'_j = \delta_0$ and C_j and C'_j are the ciphertexts corresponding to P_j and P'_j , respectively. Also, as before, for a choice κ of the target sub-key, $j = 1, \dots, N$, define $B_{\kappa,j}^{(r)}$ to be the input to the last round obtained by inverting the last round on C_j .

For the correct choice of κ , define p_i , $i = 1, \dots, v$, to be the probability that $B_{\kappa,j}^{(r)} \oplus B_{\kappa,j}^{(r)'} = \delta_r^{(i)}$. This probability is independent of both j and κ . For an incorrect choice κ , the probability that $B_{\kappa,j}^{(r)} \oplus B_{\kappa,j}^{(r)'}$ equals $\delta_r^{(i)}$ is $1/(2^n - 1)$. Define

$$p_0 = 1 - \sum_{i=1}^v p_i \quad \text{and} \quad \theta_0 = 1 - \frac{v}{2^n - 1}.$$

For the correct choice of κ , p_0 represents the probability that none of the events $B_{\kappa,j}^{(r)} \oplus B_{\kappa,j}^{(r)'} = \delta_r^{(i)}$ occur; similarly, for an incorrect choice of κ , θ_0 represents the probability that none of the events $B_{\kappa,j}^{(r)} \oplus B_{\kappa,j}^{(r)'} = \delta_r^{(i)}$ occur. So, $(p_0, p_1, \dots, p_v)$ and $(\theta_0, 1/(2^n - 1), \dots, 1/(2^n - 1))$ are both probability distributions corresponding respectively to the correct and the incorrect choices of the target sub-key. For convenience of notation we assume $\theta_i = 1/(2^n - 1)$ for $i = 1, \dots, v$ and $\tilde{p} = (p_0, p_1, \dots, p_v)$.

7.3 The LLR test statistic

For a choice κ of the target sub-key, define

$$X_{\kappa,i} = \#\{j : B_{\kappa,j}^{(r)} \oplus B_{\kappa,j}^{(r)'} = \delta_r^{(i)}\}, \quad T_\kappa = \sum_{i=0}^v X_{\kappa,i} \ln\left(\frac{p_i}{\theta_i}\right).$$

This T_κ is the LLR test statistic that was considered by Blondeau, Gérard and Nyberg in [11]. Further, using the approximation in [2], $(T_\kappa - N\mu)/(\sigma\sqrt{N})$ follows the standard normal distribution, where

$$\mu = \begin{cases} \mu_0 = \sum_{i=0}^v p_i \ln(p_i/\theta_i) & \text{if } H_0 \text{ holds,} \\ \mu_1 = \sum_{i=0}^v \theta_i \ln(\theta_i/p_i) & \text{if } H_1 \text{ holds,} \end{cases}$$

$$\sigma^2 = \begin{cases} \sigma_0^2 = \sum_{i=0}^v p_i (\ln(p_i/\theta_i))^2 - \mu_0^2 & \text{if } H_0 \text{ holds,} \\ \sigma_1^2 = \sum_{i=0}^v \theta_i (\ln(\theta_i/p_i))^2 - \mu_1^2 & \text{if } H_1 \text{ holds.} \end{cases}$$

The error analysis of the approximation is considered later. Assume that $\mu_0 > \mu_1$, the other case being handled in a similar manner. The hypothesis testing framework of (5.1) is applied with T_κ as the test statistic. Then

$$\alpha = \Pr[\text{Type-I error}] = \Pr[T_\kappa \leq t \mid H_0 \text{ holds}] = \Phi\left(\frac{t - N\mu_0}{\sigma_0\sqrt{N}}\right) = 1 - P_S,$$

$$\beta = \Pr[T_\kappa > t \mid H_1 \text{ holds}] = 1 - \Phi\left(\frac{t - N\mu_1}{\sigma_1\sqrt{N}}\right).$$

Eliminating t between these two expressions and using (5.2) to substitute 2^{-a} for β gives the following expression for N :

$$N = \frac{\{\sigma_1\Phi^{-1}(1 - 2^{-a}) + \sigma_0\Phi^{-1}(P_S)\}^2}{(\mu_0 - \mu_1)^2}.$$

This expression is identical to the expression found by Blondeau, Gérard and Nyberg in [11] using the LLR test statistics following the order statistics framework.

Remark 7.1. Due to reasons similar to those discussed in Remark 6.1, in this case also it is required for P_S to be greater than 0.5 for the analysis to be meaningful.

7.4 The χ^2 test statistic

As in [11], the test statistic is defined to be

$$T_K = N \sum_{i=0}^v \frac{(X_{K,i}/N - \theta_i)^2}{\theta_i}.$$

The hypothesis testing framework of (5.1) is applied.

From Section 6.2.1, we get that under H_0 , $T_K \sim \chi_v^2(NC(\tilde{p}))$ and under H_1 , $T_K \sim \chi_v^2$ for “sufficiently large” N and the error in approximation is considered later. Further, for large values of N , $\chi_v^2(NC(\tilde{p}))$ has been approximated by $\mathcal{N}(v + NC(\tilde{p}), 2(v + 2NC(\tilde{p})))$. Hence, under H_0 , we have

$$T_K \sim \mathcal{N}(v + NC(\tilde{p}), 2(v + 2NC(\tilde{p}))).$$

Given these distributions, the rest of the analysis is similar to that done in Section 6.2.1 and provides the following expression for the data complexity:

$$N = \frac{1}{C(\tilde{p})} \left[2(\Phi^{-1}(P_S))^2 + \Psi^{-1}(1 - 2^{-a}) - v + \Phi^{-1}(P_S) \sqrt{4(\Phi^{-1}(P_S))^2 + 2\Psi^{-1}(1 - 2^{-a})} \right],$$

where Ψ is the cumulative distribution function of the χ^2 distribution. This expression is identical to the expression found by Blondeau, Gérard and Nyberg in [11] using the χ^2 test statistics following the order statistics framework.

8 On normal approximation of the χ^2 test statistics

In deriving data complexity expressions using the hypothesis testing framework, we have made use of the earlier proposed χ^2 test statistics from [11, 20] and the corresponding normal approximations. In [20], the use was justified by a reference to the work by Drost et al. [15]. On checking [15] we found that the treatment is in terms of multinomial distribution and the error analysis is asymptotic. This did not appear to be amenable for interpreting the error analysis to the cryptanalytic context. Hence, keeping in tune with the main theme of this work, we performed a detailed analysis of the various approximations that are required. Our approach avoids the multinomial distribution. Later, we comment on how one may proceed using the multinomial distribution and argue that this would not provide better error bounds than those we obtain.

8.1 From Bernoulli to normal to χ^2 to normal

Let $\mathcal{X}$ be a set of size k and $\tilde{p} = (p_1, p_2, \dots, p_k)$ be a probability distribution on $\mathcal{X}$. Define

$$p_{\min} = \min_{\eta} p_{\eta}.$$

Note that $p_{\min} \leq 1/k$ and if $\tilde{p}$ is not the uniform distribution over $\mathcal{X}$, then $p_{\min} < 1/k$.

Let $X_1, X_2, \dots, X_{\lambda}$ be independent random variables each following $\tilde{p}$. We next describe a sequence of transformations on the random variables X_j giving rise to different random variables following different distributions.

Bernoulli distributed random variables. For $1 \leq \eta \leq k$ and $1 \leq j \leq \lambda$ define

$$Q_{\eta,j} = \begin{cases} 1 & \text{if } X_j = \eta, \\ 0 & \text{otherwise.} \end{cases}$$

Then for a fixed η , the random variables $Q_{\eta,1}, \dots, Q_{\eta,\lambda}$ are independently and identically distributed according to Bernoulli (p_{η}).

From Bernoulli to normal. For $1 \leq \eta \leq k$, define

$$Q_\eta = \sum_{j=1}^{\lambda} Q_{\eta,j}. \quad (8.1)$$

Note that $Q_\eta = \#\{j : X_j = \eta\}$ and $\sum_{\eta=1}^k Q_\eta = N$. The distribution of Q_η is to be approximated using a normal distribution. Corollary 2.2 provides the condition on λ for which such approximation can be considered to be good. Denote the bound on the right-hand side of Corollary 2.2 by Λ_η . This depends on η and it is necessary to consider the maximum of all Λ_η . Define

$$\Lambda(\tilde{p}) := \max_{\eta} \Lambda_\eta, \quad \Lambda_\eta = \frac{C(p_\eta^2 + (1 - p_\eta)^2)}{(p_\eta(1 - p_\eta))^{1/2}}.$$

The requirement is that the distributions of *all* Q_η should be approximated by appropriate normal distributions. The condition $\Lambda(\tilde{p})/\lambda^{1/2} \leq 2^{-\varepsilon}$ ensures that the errors of all approximations are at most $2^{-\varepsilon}$. So, the condition required for the errors of approximations of all Q_η by their respective normal distributions is the following:

Condition for Ber-to-Nor. $\lambda \geq 2^{2\varepsilon} \Lambda(\tilde{p})$.

From normal to χ^2 . Suppose that the condition Ber-to-Nor is satisfied. Then for each η , Q_η can be assumed to follow $\mathcal{N}(\lambda p_\eta, \lambda p_\eta(1 - p_\eta))$. There is a single constraint on the Q_η , namely $\sum_{\eta} Q_\eta = \lambda$ and so one proceeds with the assumption that $Q_1, \dots, Q_{k-1}$ are independent and there is one dependent random variable Q_k given by $Q_k = \lambda - \sum_{\eta=1}^{k-1} Q_\eta$. Let

$$Y = \sum_{\eta=1}^k \frac{(Q_\eta - \lambda p_\eta)^2}{\lambda p_\eta(1 - p_\eta)} = \lambda \times \sum_{\eta=1}^k \frac{(Q_\eta/\lambda - p_\eta)^2}{p_\eta(1 - p_\eta)}, \quad (8.2)$$

where $\tilde{\pi} = (\pi_1, \pi_2, \dots, \pi_k)$ is another distribution over $\mathcal{X}$. Two cases may arise giving rise to central or non-central χ^2 distributions.

Central χ^2 . Suppose $\tilde{p} = \tilde{\pi}$. In this case, it is well known (see [21, Chapter 18]) that Y follows the central χ^2 distribution with $k - 1$ degrees of freedom.

Non-central χ^2 . Suppose $\tilde{p} \neq \tilde{\pi}$. In this case, we would like to say that Y follows a non-central χ^2 distribution, but this requires an approximation. Write

$$\begin{aligned} Y &= \sum_{\eta=1}^k \frac{(Q_\eta - \lambda p_\eta)^2}{\lambda p_\eta(1 - p_\eta)} = \sum_{\eta=1}^k \frac{(Q_\eta - \lambda \pi_\eta)^2}{\lambda p_\eta(1 - p_\eta)} \times \left\{ 1 - \frac{p_\eta(1 - p_\eta) - \pi_\eta(1 - \pi_\eta)}{p_\eta(1 - p_\eta)} \right\}^{-1} \\ &= \sum_{\eta=1}^k \frac{(Q_\eta - \lambda \pi_\eta)^2}{\lambda p_\eta(1 - p_\eta)} \times \left\{ 1 + \frac{p_\eta(1 - p_\eta) - \pi_\eta(1 - \pi_\eta)}{p_\eta(1 - p_\eta)} + \left(\frac{p_\eta(1 - p_\eta) - \pi_\eta(1 - \pi_\eta)}{p_\eta(1 - p_\eta)} \right)^2 + \dots \right\} \\ &\approx \sum_{\eta=1}^k \frac{(Q_\eta - \lambda \pi_\eta)^2}{\lambda p_\eta(1 - p_\eta)}. \end{aligned} \quad (8.3)$$

The above computation and approximation require two conditions. The Taylor series expansion requires the condition

$$|p_\eta(1 - p_\eta) - \pi_\eta(1 - \pi_\eta)| < p_\eta(1 - p_\eta)$$

and the approximation requires

$$\frac{p_\eta(1 - p_\eta) - \pi_\eta(1 - \pi_\eta)}{p_\eta(1 - p_\eta)} \approx 0.$$

The second condition implies the first, but not the other way around. Assuming that the approximations are appropriate, the distribution of Y is given by a non-central χ^2 distribution (see [22, Section 29.1]):

$$Y \sim \chi_{k-1}^2(\lambda D_{\tilde{p}, \tilde{\pi}}), \quad D_{\tilde{p}, \tilde{\pi}} = \sum_{\eta=1}^k \frac{(p_\eta - \pi_\eta)^2}{p_\eta(1 - p_\eta)}. \quad (8.4)$$

The denominator in the non-centrality parameter is $p_\eta(1 - p_\eta)$. Using a Taylor series expansion as in the above computation, it is possible to show that

$$D_{\bar{p}, \bar{\pi}} \approx C_{\bar{p}, \bar{\pi}}, \quad (8.5)$$

where

$$C_{\bar{p}, \bar{\pi}} = \sum_{\eta=1}^k \frac{(p_\eta - \pi_\eta)^2}{\pi_\eta(1 - \pi_\eta)}.$$

The conditions required for the Taylor series expansion and the approximation (8.5) to hold are

$$|p_\eta(1 - p_\eta) - \pi_\eta(1 - \pi_\eta)| < \pi_\eta(1 - \pi_\eta) \quad \text{and} \quad \frac{p_\eta(1 - p_\eta) - \pi_\eta(1 - \pi_\eta)}{\pi_\eta(1 - \pi_\eta)} \approx 0.$$

All the required conditions for the approximations (8.3) and (8.5) hold if $\pi_\eta(1 - \pi_\eta) \approx p_\eta(1 - p_\eta)$. We record this condition.

Condition for non-central χ^2 . $\pi_\eta(1 - \pi_\eta) \approx p_\eta(1 - p_\eta)$.

From (non-central) χ^2 to normal. Consider the random variable Y following the χ^2 distribution in (8.4) under the approximation given by (8.5). Define $k - 1$ independent and identically distributed random variables Y_i such that

$$Y_i \sim \mathcal{N}\left(\sqrt{\frac{\lambda C_{\bar{p}, \bar{\pi}}}{k-1}}, 1\right) \quad \text{and} \quad Y_i^2 \sim \chi_1^2\left(\frac{\lambda C_{\bar{p}, \bar{\pi}}}{k-1}\right), \quad i = 1, 2, \dots, k-1$$

(see “Characterization” in [22, Section 29.5]). It is known (see “Reproductivity” in [22, Section 29.5]) that $\sum_{i=1}^{k-1} Y_i^2$ follows $\chi_{k-1}^2(\lambda C_{\bar{p}, \bar{\pi}})$ which is also the distribution of Y . So, to approximate the distribution of Y by a normal, it is sufficient to approximate the distribution of $\sum_{i=1}^{k-1} Y_i^2$ by normal. We show that the Berry–Esséen theorem applies and provides the error in approximation.

For each $i = 1, 2, \dots, k-1$ we have

$$Y_i^2 \sim \chi_1^2\left(\frac{\lambda C_{\bar{p}, \bar{\pi}}}{k-1}\right) \quad \text{and} \quad E[Y_i^2] = 1 + \frac{\lambda C_{\bar{p}, \bar{\pi}}}{k-1}.$$

Let

$$Z_i = Y_i^2 - \left(1 + \frac{\lambda C_{\bar{p}, \bar{\pi}}}{k-1}\right).$$

Then, $E[Z_i] = 0$,

$$\sigma^2 = E[Z_i^2] = 2\left(1 + 2\frac{\lambda C_{\bar{p}, \bar{\pi}}}{k-1}\right) = \frac{2}{k-1}(k-1 + 2\lambda C_{\bar{p}, \bar{\pi}}),$$

and

$$\begin{aligned} \rho &= E[|Z_i|^3] \leq E[(Z_i^2)^3] + \left(1 + \frac{\lambda C_{\bar{p}, \bar{\pi}}}{k-1}\right)^3 \\ &= 2\left(1 + \frac{\lambda C_{\bar{p}, \bar{\pi}}}{k-1}\right)^3 + 6\left(1 + \frac{\lambda C_{\bar{p}, \bar{\pi}}}{k-1}\right)\left(1 + 2\frac{\lambda C_{\bar{p}, \bar{\pi}}}{k-1}\right) + 8\left(1 + 3\frac{\lambda C_{\bar{p}, \bar{\pi}}}{k-1}\right) \\ &= \frac{1}{(k-1)^3} \{2(k-1 + \lambda C_{\bar{p}, \bar{\pi}})^3 + 6(k-1)(k-1 + \lambda C_{\bar{p}, \bar{\pi}})(k-1 + 2\lambda C_{\bar{p}, \bar{\pi}}) + 8(k-1)^2(k-1 + 3\lambda C_{\bar{p}, \bar{\pi}})\} \\ &= \rho' \quad (\text{say}), \end{aligned}$$

where the inequality holds since $Y_i^2 > 0$ and $1 + \lambda C_{\bar{p}, \bar{\pi}}/(k-1) > 0$.

Note that for $C_{\bar{p}, \bar{\pi}} = 0$, we have $\sigma^2 = 2$ and $\rho' = 16$. More generally, $\rho \leq \rho' < \infty$. So, Theorem 2.1 applies and we have

$$\begin{aligned} \left| \Pr\left[\frac{\sum_{i=1}^{k-1} Z_i}{\sqrt{k-1}\sigma} < x\right] - \Phi(x) \right| &= \left| \Pr\left[\frac{\sum_{i=1}^{k-1} Y_i^2 - (k-1 + \lambda C_{\bar{p}, \bar{\pi}})}{\sqrt{k-1}\sigma} < x\right] - \Phi(x) \right| \\ &= \left| \Pr\left[\frac{Y - (k-1 + \lambda C_{\bar{p}, \bar{\pi}})}{\sqrt{k-1}\sigma} < x\right] - \Phi(x) \right| \\ &\leq \frac{C\rho}{\sigma^3 \sqrt{k-1}} \leq \frac{C\rho'}{\sigma^3 \sqrt{k-1}}. \end{aligned} \quad (8.6)$$

If the bound on the right-hand side of (8.6) is at most $2^{-\varepsilon}$, then the distribution of Y can be approximated by $\mathcal{N}(k - 1 + \lambda C_{\tilde{p}, \tilde{\pi}}, 2(k - 1 + 2\lambda C_{\tilde{p}, \tilde{\pi}}))$ and the error in approximation will be at most $2^{-\varepsilon}$.

If $C_{\tilde{p}, \tilde{\pi}} = 0$, then Y follows χ^2_{k-1} which is approximated by $\mathcal{N}(k - 1, 2(k - 1))$ with error in approximation at most $2^{-\varepsilon}$ if $k - 1 \geq C^2 2^{2\varepsilon+5}$. In the more general situation, the bound in (8.6) is difficult to simplify. If, however, $C_{\tilde{p}, \tilde{\pi}}$ is “small”, then the condition $k - 1 \geq C^2 2^{2\varepsilon+5}$ again ensures that the error in approximating the distribution of Y by normal is at most $2^{-\varepsilon}$.

Condition for central χ^2 to normal. $k - 1 \geq C^2 \times 2^{2\varepsilon+5}$ (under the condition $C_{\tilde{p}, \tilde{\pi}}$ is “small”).

8.2 Application to linear cryptanalysis

We interpret the above result in the context of the χ^2 statistics as used in linear cryptanalysis [20] (see also Section 6.2.1). In this context, the set $\mathcal{X}$ is $\{0, 1\}^\ell$ and so $k = 2^\ell$, $\lambda = N$, $\tilde{p}$ is the distribution under the correct key, and $\tilde{\pi}$ is the uniform distribution over $\{0, 1\}^\ell$. Under these conditions, the random variable Y given by (8.2) is not the same as the test statistic T_κ given in (6.6). On the other hand, if we make the additional assumption that $1 - \pi_\eta \approx 1$, then Y given by (8.2) becomes identical to T_κ given in (6.6). Since $\pi_\eta = 2^{-\ell}$, the condition $1 - \pi_\eta \approx 1$ amounts to $1 - 2^{-\ell} \approx 1$ which says that the number of linear approximations ℓ must be large enough for this approach to be valid. We record this condition.

Condition for T_κ given in (6.6) to be equal to Y given by (8.2). $1 - 2^{-\ell} \approx 1$.

Assuming this approximation, we have $C_{\tilde{p}, \tilde{\pi}} = C(\tilde{p})$. The distributions of T_κ both for the correct choice of κ and an incorrect choice of κ are given by (6.7). This requires certain conditions as derived in Section 8.1. The conditions required for the approximation to normal via the non-central χ^2 distributions are a superset of those required for the path via the central χ^2 distribution. Below we interpret these conditions in the context of linear cryptanalysis.

- Condition Ber-to-Nor: For $1 \leq \eta \leq 2^\ell - 1$,

$$N \geq 2^{2\varepsilon} \times \Lambda_{\tilde{p}} = \max_{\eta} 2^{2\varepsilon} C^2 \times \frac{(p_\eta^2 + (1 - p_\eta)^2)^2}{p_\eta(1 - p_\eta)}.$$

Since the condition is to be true for all p_η , it must hold for $p_{\min}$. Noting that $p_{\min} \leq 2^{-\ell}$, we have the condition

$$N \geq 2^{2\varepsilon} C^2 \times \frac{(p_{\min}^2 + (1 - p_{\min})^2)^2}{p_{\min}(1 - p_{\min})} \approx \frac{2^{2\varepsilon} C^2}{p_{\min}} \geq \frac{2^{2\varepsilon} C^2}{2^{-\ell}} = C^2 \times 2^{2\varepsilon+\ell}.$$

This shows an unexpected fact. For the conversion of the sum of the Bernoulli variables $\sum_{j=1}^N Q_{\eta,j}$ to be well approximated by the normal, it is required for the number of plaintext-ciphertext pairs to be more than 2^ℓ . So, as the number of linear approximations grows, one needs to obtain more plaintext-ciphertext pairs just for the approximation to be valid. This is counter-intuitive, since one would expect that as ℓ increases, the data complexity N should go down.

- Condition for non-central: In this case $\tilde{\pi}$ is the uniform distribution over $\{0, 1\}^\ell$ and so this condition becomes $p_\eta(1 - p_\eta) \approx 2^{-\ell}$ for all η . This puts a condition on the distribution $\tilde{p}$ that for all η , $p_\eta(1 - p_\eta)$ should be close to $2^{-\ell}$. If a real life distribution does not satisfy this condition, then the approximation does not hold. So, if a cryptanalyst is attacking a not so well-designed block cipher and is able to uncover linear approximations where for some η , $p_\eta(1 - p_\eta)$ is significantly away from $2^{-\ell}$, then (s)he cannot apply the χ^2 test statistics to assess the data complexity. This is again counter-intuitive since having a $p_\eta \approx p_\eta(1 - p_\eta)$ to be significantly away from $2^{-\ell}$ should help rather than hinder the analysis.
- Condition for χ^2 to normal: The condition is $2^\ell - 1 \geq C^2 \times 2^{2\varepsilon+5}$. This means that the χ^2 statistics based analysis cannot be applied if ℓ is low. If $\varepsilon = 10$ then ℓ should be at least about 25. So, for a smaller ℓ , the applicability of the method is in question.

Avoiding the approximation of χ^2 by normal. The last approximations consist of approximating the χ^2 distributions (both central and non-central) by appropriate normal distributions. As explained in Section 4.1, for the order statistics based approach this is necessary. For the hypothesis testing based approach, these approximations can be avoided and one can work directly with the distribution functions of the central χ^2 (corresponding to an incorrect key) and the non-central χ^2 (corresponding to the correct key) distributions. This allows the expression of the data complexity in terms of these distribution functions.

8.3 Application to differential cryptanalysis

In this case, the k of Section 8.1 is $\ell + 1$. The observation made for the non-central χ^2 approximation in the context of linear cryptanalysis remains unchanged. The other two observations are to be modified with the new value of k . This somewhat mitigates the criticisms though the general spirit still remains valid.

We provide a concrete example to show that if the conditions for the approximations are not satisfied, then the expression for data complexity may give meaningless results.

Consider two distributions $\tilde{p} = (p_0, \dots, p_v)$ and $\tilde{q} = (q_0, \dots, q_v)$. In the context of multiple differential cryptanalysis using the chi-squared approach, as discussed earlier, the test statistic follows a non-central chi-squared distribution for the correct choice of the target sub-key. We investigate this assumption in more details in the context of a 64-bit toy block cipher SMALLPRESENT considered in [12]. Table 6 of [10] gives a list of 55 9-round differentials for this block cipher. For each differential, the table gives three probability estimates, namely, a theoretical estimate, an estimate using 40-bit key scheduling and an estimate using 80-bit key scheduling. These 55 differentials are divided into six classes where the differentials in each class have the same input difference.

In the multiple differential attack considered in this work, all the differentials should have the same input difference. So, the 55 differentials given in [10] gives rise to the possibility of 6 different (same input difference) multiple differential attacks. Taken together with the 3 different probability estimates, gives a total of 18 different distributions with which multiple differential attacks can be applied.

Following the chi-squared approach, as mentioned earlier, for the non-central chi-squared approximation to hold, it is necessary that for each $\eta = 0, 1, \dots, v$ both of the following two conditions are satisfied:

$$|p_\eta(1 - p_\eta) - q_\eta(1 - q_\eta)| \leq p_\eta(1 - p_\eta) \quad \text{and} \quad \frac{|p_\eta(1 - p_\eta) - q_\eta(1 - q_\eta)|}{p_\eta(1 - p_\eta)} \approx 0.$$

An experiment was conducted to check whether the distributions satisfy these conditions. The experiment showed that for all the 18 distributions, the first condition is satisfied, but that the left-hand side of the second condition is ≈ 1 . As a consequence, the data complexities computed from these distributions give absurd results.

The column labelled N_{χ^2} in Table 1 gives the data complexities of the chi-squared test statistic for the 18 distributions that arise from [10, Table 6]. The data complexities were calculated for $P_S = 0.5$ and $\alpha = 20$. The paper mentions that these differentials were used to recover 32 bits of the last two round sub-keys and so $m = 32$. We explain the results with the help of a particular example. Consider the first entry of the table, i.e., the theoretical estimate of the distribution corresponding to the input difference 0x3. In this case, the distributions corresponding to the correct and the incorrect choice of the target sub-key are given by

$$\tilde{p} \approx (0.99, 7.67 \times 10^{-10}, 7.41 \times 10^{-10}, 6.77 \times 10^{-10}, 6.23 \times 10^{-10}, 6.19 \times 10^{-10}, 5.98 \times 10^{-10}, 5.73 \times 10^{-10}, 5.73 \times 10^{-10}, 4.75 \times 10^{-10})$$

and

$$\tilde{q} \approx (0.99, 5.42 \times 10^{-20}, 5.42 \times 10^{-20}, 5.42 \times 10^{-20}, 5.42 \times 10^{-20}, 5.42 \times 10^{-20}, 5.42 \times 10^{-20}, 5.42 \times 10^{-20}, 5.42 \times 10^{-20}, 5.42 \times 10^{-20}),$$

respectively, where $v = 9$.

Probability estimates	Input difference	N_{χ^2}
Theoretical	0x3	0.54
	0x7	0.18
	0xD	0.28
	0x5	0.47
	0xB	0.53
	0xF	0.20
40-bit	0x3	0.24
	0x7	0.11
	0xD	0.14
	0x5	0.23
	0xB	0.23
	0xF	0.11
80-bit	0x3	0.28
	0x7	0.0002
	0xD	0.17
	0x5	0.23
	0xB	0.27
	0xF	0.12

Table 1. Table showing the data complexities of multiple differential cryptanalysis following the chi-squared test statistic using the distributions arising from [10, Table 6] with $P_S = 0.5$ and $a = 20$.

Recall that the data complexity for the chi-squared test statistic is given by

$$\begin{aligned}
 N &= \frac{2(\Phi^{-1}(P_S))^2 + \Psi_v^{-1}(1 - 2^{-a}) - v + \Phi^{-1}(P_S) \sqrt{2\Psi_v^{-1}(1 - 2^{-a}) + 4(\Phi^{-1}(P_S))^2}}{C(\tilde{p}, \tilde{q})} \\
 &= \frac{\Psi_9^{-1}(1 - 2^{-20}) - 20}{C(\tilde{p}, \tilde{q})} \quad (\text{putting } P_S = 0.5, a = 20 \text{ and } v = 9) \\
 &\approx \frac{35.92}{C(\tilde{p}, \tilde{q})},
 \end{aligned}$$

where $C(\tilde{p}, \tilde{q}) = \sum_{i=0}^v (p_i - q_i)^2 / q_i$. The value of $C(\tilde{p}, \tilde{q})$ in this case is approximately equal to 66.54. Therefore the data complexity becomes approximately equal to 0.54, which is meaningless.

The above shows a practical example where the failure of a condition for approximation leads to an meaningless value of the data complexity. This highlights the importance of analysing the conditions for an approximation to hold and to evaluate these conditions before applying the formula for data complexity.

8.4 The multinomial approach

The tuple $(Q_1, \dots, Q_k)$ defined using (8.1) follows the multinomial distribution with parameters λ and $\tilde{p}$. It is possible to approximate this distribution using a multivariate normal distribution. The error bound of such an approximation is known and given by [33, Theorem 3]. Similar to that of the Berry–Esséen theorem, this bound involves a constant $C'(k)$ which depends only on k and has been shown to be $O(k^{1/2})$ (see [4]). We have worked out the error bound following this approach and have found that the lower bound on λ that is obtained from this approach is more than what is given by the condition “Ber-to-Nor”. In view of this, we do not provide the details of this approach.

9 On normal approximation of the LLR test statistics

There are two issues to the approximation. The first approximation is to show that the distribution of the LLR test statistics can be approximated by a normal distribution. The parameters of the normal distribution are in terms of the Kullback–Leibler divergence of two distributions. The second approximation is to show that the

Kullback–Leibler divergence can be approximated by functions of the capacity. The outline of how the two approximations can be carried out was already described in [2, 20]. Here we apply the Berry–Esséen theorem to work out the error bound for the first approximation. Obtaining a closed form expression for this error seems to be difficult which makes it difficult to deduce the conditions under which the approximation is good. The second approximation uses the Taylor series expansion of the logarithmic function. The condition for such an approximation to hold yields an unexpected consequence. We further argue that this approximation is, in fact, not required.

9.1 Normal approximation

Recall from (6.12) the LLR test statistics

$$T_{\kappa,z} = \sum_{\eta \in \{0,1\}^\ell} Q_{\kappa,z,\eta} \ln \left(\frac{\tilde{p}_z(\eta)}{2^{-\ell}} \right).$$

This can be written in terms of $X_{\kappa,z,j}$ for $j = 1, \dots, N$ as follows:

$$T_{\kappa,z} = \sum_{\eta \in \{0,1\}^\ell} Q_{\kappa,z,\eta} \times \ln \left(\frac{\tilde{p}_z(\eta)}{2^{-\ell}} \right) = \sum_{j=1}^N \ln \left(\frac{\tilde{p}_z(X_{\kappa,z,j})}{2^{-\ell}} \right).$$

Recall that for a fixed κ and z , the random variables $X_{\kappa,z,j}$ are independent and identically distributed random variables. For $j = 1, \dots, N$, let $Y_j = \ln(\tilde{p}_z(X_{\kappa,z,j})/2^{-\ell})$. Then $T_{\kappa,z} = \sum_{j=1}^N Y_j$ where the random variables Y_j are independent and identically distributed random variables.

Under H_i ($i = 0, 1$), we have $\mu_i = E[Y_j]$, $\sigma_i^2 = E[(Y_j - \mu_i)^2]$ and let $\rho_i = E[|Y_j - \mu_i|^3]$. Since the random variables Y_j take values from a finite set, clearly $\rho_i < \infty$ and so Theorem 2.1 applies and $T_{\kappa,z} = \sum_{j=1}^N Y_j$ approximately follows a normal distribution. The error in the approximation is upper bounded by $C\rho_i/(\sigma_i^3 N^{1/2})$. For this upper bound to be at most $2^{-\varepsilon}$, N has to satisfy

$$N \geq \max \left(\frac{C^2 \rho_0^3 2^{2\varepsilon}}{\sigma_0^6}, \frac{C^2 \rho_1^3 2^{2\varepsilon}}{\sigma_1^6} \right).$$

This gives a lower bound on N which needs to be contrasted with the expression for N given by (6.16) and (6.17).

This analysis has been done in the context of multiple linear cryptanalysis. A similar analysis holds analogously for multiple differential cryptanalysis.

9.2 Taylor series expansion

Baignères, Junod and Vaudenay [2] mentioned that using the Taylor series expansion of the natural logarithm function, it is possible to obtain approximate expressions for μ_0, μ_1, σ_0 and σ_1 . This discussion is applicable only for multiple linear cryptanalysis as the approximations have not been suggested for multiple differential cryptanalysis.

The approximations were suggested if all ϵ_η are small. In that case, it was mentioned that $T_{\kappa,z}$ asymptotically follows $\mathcal{N}(N\mu, N\sigma^2)$, where

$$\mu = \begin{cases} \mu_0 = D(\tilde{p}_z \parallel \tilde{p}_\$) \approx \frac{1}{2} C(\tilde{p}) & \text{if } H_0 \text{ holds,} \\ \mu_1 = -D(\tilde{p}_\$ \parallel \tilde{p}_z) \approx -\frac{1}{2} C(\tilde{p}) & \text{if } H_1 \text{ holds,} \end{cases} \quad \text{and} \quad \sigma^2 \approx C(\tilde{p}).$$

We consider further details of this approximation:

$$D(\tilde{p}_z \parallel \tilde{p}_\$) = \sum_{\eta \in \{0,1\}^m} \tilde{p}_z(\eta) \ln \frac{\tilde{p}_z(\eta)}{\tilde{p}_\$(\eta)} = \sum_{\eta \in \{0,1\}^m} \tilde{p}_z(\eta) \ln \left(\frac{2^{-\ell} + \epsilon_{\eta \oplus z}}{2^{-\ell}} \right) = \sum_{\eta \in \{0,1\}^m} \tilde{p}_z(\eta) \ln \left(1 + \frac{\epsilon_{\eta \oplus z}}{2^{-\ell}} \right).$$

For each η , the Taylor series expansion of $\ln(1 + \epsilon_{\eta \oplus z}/2^{-\ell})$ is made and up to quadratic terms are retained while cubic and higher order terms are discarded. From this a routine calculation shows the desired expressions in terms of $C(\tilde{p})$.

The important point to note is that the expansions are valid only if for each $\epsilon_{\eta \oplus z}$ the following holds:

$$\left| \frac{\epsilon_{\eta \oplus z}}{2^{-\ell}} \right| < 1.$$

This in particular means that $|\epsilon_{\eta}| < 2^{-\ell}$ for all η . The range of ϵ_{η} is $-2^{-\ell} \leq \epsilon_{\eta} \leq 1 - 2^{-\ell}$. So, even if for one η it happens that $2^{-\ell} \leq \epsilon_{\eta} \leq 1 - 2^{-\ell}$, then the Taylor series expansion cannot be performed for this ϵ_{η} and so the overall approximations of the means and the variances in terms of the capacity do not hold. For such a situation, the data complexity given by (6.17) is not valid.

Consider for example that $\epsilon_{\eta} = 2^{-\ell}$ for some η . Then the corresponding probability is $2^{-\ell+1}$ which is twice the probability that would be expected if the distribution were uniform. Such a situation can indeed occur in practice. In the context of stream ciphers, Mantin and Shamir [28] showed that the probability that the second output byte of the RC4 stream cipher is 0 is twice of what would be expected from a uniform distribution. Although this is an example from the domain of stream ciphers, this does point to the possibility that even in the context of block ciphers such a situation may arise. The expression for data complexity given by (6.17) cannot be applied to this case.

Redundancy of the Taylor series expansion. The expression for data complexity given by (6.16) does not require the Taylor series expansion. One can take this as the data complexity given by the LLR based test statistics. There are apparently two problems with doing this. First, the expressions for μ_0, μ_1, σ_0 and σ_1 given by (6.13) and (6.14) appear to depend on z and so the data complexity may also appear to depend on z . Second, the data complexity expression given in terms of the actual values of μ_0, μ_1, σ_0 and σ_1 is not compact and hence is less useful. We show that none of these are valid issues.

Consider $D(\tilde{p}_z \| \tilde{p}_{\S})$. We show that this value does not depend on z . Note that $\tilde{p}_{\S}$ is the uniform distribution on $\{0, 1\}^{\ell}$ and so does not depend on z . For any $\beta \in \{0, 1\}^{\ell}$ consider $D(\tilde{p}_{z \oplus \beta} \| \tilde{p}_{\S})$:

$$\begin{aligned} D(\tilde{p}_{z \oplus \beta} \| \tilde{p}_{\S}) &= \sum_{\eta} \tilde{p}_{z \oplus \beta}(\eta) \ln \left(\frac{\tilde{p}_{z \oplus \beta}(\eta)}{\tilde{p}_{\S}(\eta)} \right) \\ &= \sum_{\eta} \tilde{p}_z(\eta \oplus \beta) \ln \left(\frac{\tilde{p}_{z \oplus \beta}(\eta)}{\tilde{p}_{\S}(\eta)} \right) \\ &= \sum_{\eta} \tilde{p}_z(\eta) \ln \left(\frac{\tilde{p}_{z \oplus \beta}(\eta)}{\tilde{p}_{\S}(\eta)} \right). \end{aligned}$$

Here we have used (3.2) in the first step. Working from the last expression, one can show that the expressions for μ_0, μ_1, σ_0 and σ_1 are given as follows:

$$\mu = \begin{cases} \mu_0 = D(\tilde{p} \| \tilde{p}_{\S}) & \text{if } H_0 \text{ holds,} \\ \mu_1 = -D(\tilde{p}_{\S} \| \tilde{p}) & \text{if } H_1 \text{ holds,} \end{cases} \quad (9.1)$$

$$\sigma^2 = \begin{cases} \sigma_0^2 = \sum_{\eta=1}^{\nu} p(\eta) (\ln(\frac{p(\eta)}{p_{\S}(\eta)}))^2 - \mu_0^2 & \text{if } H_0 \text{ holds,} \\ \sigma_1^2 = \sum_{\eta=1}^{\nu} p_{\S}(\eta) (\ln(\frac{p_{\S}(\eta)}{p(\eta)}))^2 - \mu_1^2 & \text{if } H_1 \text{ holds.} \end{cases} \quad (9.2)$$

The distribution of $\tilde{p}$ is as defined by (3.3). Using these values in (6.16) gives an expression for data complexity which is independent of z .

Next we consider the issue of whether the expression for data complexity given by (6.16) in terms of (9.1) and (9.2) is less useful than the expression for data complexity given by (6.17). To evaluate (6.17), it is required to evaluate $C(\tilde{p})$ which requires $O(2^{\ell})$ time. Evaluating (6.16), on the other hand, requires evaluating (9.1) and (9.2) which also requires $O(2^{\ell})$ time. So, evaluating both (6.17) and (6.16) requires $O(2^{\ell})$ time. Hence, in terms of being able to compute there is no advantage in considering (6.17) over (6.16). From the point of view of applicability, however, using (6.16) avoids the unnecessary restriction imposed by the Taylor series expansions required to obtain (6.17).

9.3 Non-applicability for $P_S \leq 0.5$

Recall that Remarks 6.1 and 7.1 mention that it is required for P_S to be greater than 0.5 for the analysis to be meaningful. For $P_S \leq 0.5$, the analysis and hence the expression for data complexity is not applicable. In fact, experiments show that such computations result in meaningless values of the data complexity. The reason for the condition $P_S > 0.5$ arises due to the normal approximation of the distribution of the LLR test statistic. This issue has been mentioned earlier in [20, Theorem 2] in the context of multiple linear cryptanalysis. No explanations were, however, provided. In the context of differential cryptanalysis, the LLR based test statistic was used to derive expression for data complexity in [11]. The condition $P_S > 0.5$ is also required for this analysis to be meaningful, though the condition was not mentioned in [11].

The condition $P_S > 0.5$ is restrictive. From a cryptanalytic point of view, an attack can be considered successful even if the success probability is less than 0.5. For example, an attack with success probability 0.1 will be able to recover the correct key in 10% of cases. Such an attack should be considered to have broken a block cipher. On the other hand, for such an attack, the normal approximation based approach will not be able to estimate the data complexity. In other words, it will not be possible to say how much data is required for an attack to be successful in 10% of the cases. In our opinion, this is a serious drawback of the normal approximation approach.

10 Conclusion

Normal approximations of the distribution of test statistics are often used in the statistical analysis of attacks on symmetric key ciphers. Such approximations are mostly justified in an asymptotic sense. In this work, we study the issue of normal approximations in greater depth. In concrete terms, the error in such an approximation can be bounded by the Berry–Esséen theorem. Using this as a basic tool, we carefully consider some of the important normal approximation results used earlier. These include the normal approximation of the distribution of order statistics, the χ^2 test statistics and the LLR test statistics.

The results that we obtained are surprising. For one thing, the framework of order statistics based approach to key recovery attacks turns out to have limited applicability. We showed that previous results on data complexities obtained using the order statistics based approach can be obtained using the simpler hypothesis testing based framework. The later approach avoids the difficulties that arise in the former approach.

We also carefully considered the issues related to the approximations involved in using the χ^2 and the LLR test statistics. It turns out that these have some counter-intuitive consequences which limit their applicability. All the issues that our work uncovered could not be satisfactorily resolved and resolving these can form the motivation for future works. Normal approximations are used in more cryptanalytic contexts than what could be considered in this work. It would be of interest to work out the concrete error bounds and their consequences. This is true not only of normal approximations but for any other approximations used in cryptanalysis.

References

- [1] M. A. Abdelraheem, M. Ågren, P. Beelen and G. Leander, On the distribution of linear biases: Three instructive examples, in: *Advances in Cryptology* (CRYPTO 2012), Lecture Notes in Comput. Sci. 7417, Springer, Berlin (2012), 50–67.
- [2] T. Baignères, P. Junod and S. Vaudenay, How far can we go beyond linear cryptanalysis?, in: *Advances in Cryptology* (ASIACRYPT 2004), Lecture Notes in Comput. Sci. 3329, Springer, Berlin (2004), 432–450.
- [3] T. Baignères, P. Sepehrdad and S. Vaudenay, Distinguishing distributions using Chernoff information, in: *Provable Security*, Lecture Notes in Comput. Sci. 6402, Springer, Berlin (2010), 144–165.
- [4] V. Bentkus, Dependence of the Berry–Esséen estimate on the dimension, *Lithuanian Math. J.* **26** (1986), 110–114.
- [5] A. C. Berry, The accuracy of the Gaussian approximation to the sum of independent variates, *Trans. Amer. Math. Soc.* **49** (1941), 122–136.

- [6] E. Biham and A. Shamir, Differential cryptanalysis of DES-like cryptosystems, in: *Advances in Cryptology* (CRYPTO'90), Lecture Notes in Comput. Sci. 537, Springer, Berlin (1990), 2–21.
- [7] E. Biham and A. Shamir, Differential cryptanalysis of DES-like cryptosystems, *J. Cryptology* **4** (1991), 3–72.
- [8] A. Biryukov, C. De Cannière and M. Quisquater, On multiple linear approximations, in: *Advances in Cryptology* (CRYPTO 2004), Lecture Notes in Comput. Sci. 3152, Springer, Berlin (2004), 1–22.
- [9] C. Blondeau, A. Bogdanov and G. Leander, Bounds in shallows and in miseries, in: *Advances in Cryptology* (CRYPTO 2013), Lecture Notes in Comput. Sci. 8042, Springer, Berlin (2013), 204–221.
- [10] C. Blondeau and B. Gérard, Multiple differential cryptanalysis: Theory and practice, in: *Fast Software Encryption*, Lecture Notes in Comput. Sci. 6733, Springer, Berlin (2011), 35–54.
- [11] C. Blondeau, B. Gérard and K. Nyberg, Multiple differential cryptanalysis using LLR and χ^2 statistics, in: *Security and Cryptography for Networks*, Lecture Notes in Comput. Sci. 7485, Springer, Berlin (2012), 343–360.
- [12] C. Blondeau, B. Gérard and J.-P. Tillich, Accurate estimates of the data complexity and success probability for various cryptanalyses, *Des. Codes Cryptogr.* **59** (2011), 3–34.
- [13] A. Bogdanov and E. Tischhauser, On the wrong key randomisation and key equivalence hypotheses in Matsui's algorithm 2, in: *Fast Software Encryption*, Lecture Notes in Comput. Sci. 8424, Springer, Berlin (2014), 19–38.
- [14] J. Daemen and V. Rijmen, Probability distributions of correlation and differentials in block ciphers, *J. Math. Crypt.* **1** (2007), 221–242.
- [15] F. C. Drost, W. C. M. Kallenberg, D. S. Moore and J. Oosterhoff, Power approximations to multinomial tests of fit, *J. Amer. Statist. Assoc.* **84** (1989), 130–141.
- [16] C.-G. Esséen, On the Liapounoff limit of error in the theory of probability, *Ark. Mat. Astron. Fys.* **A28** (1942), 1–19.
- [17] C.-G. Esséen, A moment inequality with an application to the central limit theorem, *Scand. Actuar. J.* **1956** (1956), 160–170.
- [18] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, John Wiley & Sons, New York, 2008.
- [19] C. Harpes, G. G. Kramer and J. L. Massey, A generalization of linear cryptanalysis and the applicability of Matsui's piling-up lemma, in: *Advances in Cryptology* (EUROCRYPT'95), Lecture Notes in Comput. Sci. 921, Springer, Berlin (1995), 24–38.
- [20] M. Hermelin, J. Y. Cho and K. Nyberg, Multidimensional extension of Matsui's algorithm 2, in: *Fast Software Encryption*, Lecture Notes in Comput. Sci. 5665, Springer, Berlin (2009), 209–227.
- [21] N. L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions*, Vol. 1, 2nd ed., John Wiley & Sons, New York, 1994.
- [22] N. L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions*, Vol. 2, 2nd ed., John Wiley & Sons, New York, 1995.
- [23] P. Junod, On the optimality of linear, differential, and sequential distinguishers, in: *Advances in Cryptology* (EUROCRYPT 2003), Lecture Notes in Comput. Sci. 2656, Springer, Berlin (2003), 17–32.
- [24] P. Junod and S. Vaudenay, Optimal key ranking procedures in a statistical cryptanalysis, in: *Fast Software Encryption*, Lecture Notes in Comput. Sci. 2887, Springer, Berlin (2003), 235–246.
- [25] B. S. Kaliski Jr and M. J. B. Robshaw, Linear cryptanalysis using multiple approximations, in: *Advances in Cryptology* (Crypto'94), Lecture Notes in Comput. Sci. 839, Springer, Berlin (1994), 26–39.
- [26] L. R. Knudsen, Truncated and higher order differentials, in: *Fast Software Encryption*, Lecture Notes in Comput. Sci. 1008, Springer, Berlin (1995), 196–211.
- [27] G. Leander, On linear hulls, statistical saturation attacks, PRESENT and a cryptanalysis of PUFFIN, in: *Advances in Cryptology* (EUROCRYPT 2011), Lecture Notes in Comput. Sci. 6632, Springer, Berlin (2011), 303–322.
- [28] I. Mantin and A. Shamir, A practical attack on broadcast RC4, in: *Fast Software Encryption*, Lecture Notes in Comput. Sci. 2355, Springer, Berlin (2002), 152–164.
- [29] M. Matsui, Linear cryptanalysis method for DES cipher, in: *Advances in Cryptology* (EUROCRYPT'93), Lecture Notes in Comput. Sci. 765, Springer, Berlin (1993), 386–397.
- [30] M. Matsui, The first experimental cryptanalysis of the data encryption standard, in: *Advances in Cryptology* (CRYPTO'94), Lecture Notes in Comput. Sci. 839, Springer, Berlin (1994), 1–11.
- [31] S. Murphy, The independence of linear approximations in symmetric cryptanalysis, *IEEE Trans. Inform. Theory* **52** (2006), 5510–5518.
- [32] S. Murphy, F. Piper, M. Walker and P. Wild, Likelihood estimation for block cipher keys, Technical Report RHUL-MA-2006-3, Royal Holloway, University of London, 1995.
- [33] V. V. Sazonov, On the multi-dimensional central limit theorem, *Sankhya A* **30** (1968), 181–204.
- [34] A. A. Selçuk, On probability of success in linear and differential cryptanalysis, *J. Cryptology* **21** (2008), 131–147.
- [35] A. Tardy-Corffdir and H. Gilbert, A known plaintext attack of FEAL-4 and FEAL-6, in: *Advances in Cryptology* (CRYPTO'91), Lecture Notes in Comput. Sci. 576, Springer, Berlin (1991), 172–181.
- [36] I. S. Tyurin, An improvement of upper estimates of the constants in the Lyapunov theorem, *Russian Math. Surveys* **65** (2010), 201–202.
- [37] A. M. Walker, A note on the asymptotic distribution of sample quantiles, *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **30** (1968), 570–575.