Another Look at Normal Approximations in Cryptanalysis

Subhabrata Samajder and Palash Sarkar
Applied Statistics Unit
Indian Statistical Institute
203, B.T.Road, Kolkata, India - 700108.
{subhabrata_r,palash}@isical.ac.in

June 28, 2015

Abstract

Statistical analysis of attacks on symmetric ciphers often require 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. Analysing these error bounds in the cryptanalytic context throws up 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. Detailed consideration of the error in normal approximation also has implications for χ^2 and the log-likelihood ratio (LLR) based test statistics. The normal approximation of the χ^2 test statistics has some serious and counter-intuitive restrictions. One such restriction is that for multiple linear cryptanalysis as the number of linear approximations grows so does the requirement on the number of plaintext-ciphertext pairs for the approximation to be proper. The issue of satisfactorily addressing the problems with the application of the χ^2 test statistics remains open. Normal approximation of the LLR test statistics too has some problematic issues which limit its applicability. 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 cryptanalysis, differential cryptanalysis, χ^2 test, log-likelihood test, order statistics, normal distribution.

1 Intoduction

Let $X_1, \ldots, X_\lambda$ be independent and identically distributed random variables having mean μ and variance σ^2 and taking values from a finite non-empty set. Asymptotically, $((X_1 + \cdots + X_\lambda)/\lambda - \mu)/(\sigma\sqrt{\lambda})$ follows the standard normal distribution as $\lambda \to \infty$. 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 X_i's follow $\mathsf{Bernoulli}(p)$ 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 papers in the literature 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-Esséen 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 security and concrete security 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. Our findings throw up some very surprising and counter-intuitive facts which invalidates several previous works. In some cases, we are able to recover the results following a different technique while in other cases, providing satisfactory replacements seem 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. 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 are 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, P_S, m and possibly 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 [33] had 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 had 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 et al [20] for multiple linear cryptanalysis and by Blondeau et al [11] for multiple differential cryptanalysis. Both these works had obtained expressions for data complexities. The criticisms of the order statistics based approach puts into doubt the applicability of the data complexity results from [33, 20, 11].

Hypothesis testing based framework for key recovery attacks: We show that the data complexity results from [33, 20, 11] can be recovered by following the hypothesis testing framework for analysing key recovery attacks. This framework appears in the literature in the context of analysing distinguishing attacks and does not seem to have been seriously considered for analysing key recovery attacks. We find this to be surprising, since the hypothesis testing framework is simpler and avoids the restrictions of the order statistics based approach. In re-deriving the results of [20, 11], 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 [20, 11] requires several approximations. Carefully analysing these show some very surprising results. We mention these as applied to multiple linear cryptanalysis. These also hold for multiple differential cryptanalysis, though to a lesser extent.

1. 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.

2. 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.

3. The joint distribution of the ℓ linear approximations must be close to the uniform distribution at every point. This is surprising, 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. To the best of our knowledge, no detailed treatment of these issues and their implications appear earlier. On the other hand, even though we are able to uncover these issues, the problem of satisfactorily addressing these issues remain open.

The LLR based test statistics used in [20, 11] 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 LLR based approach may not be applicable in certain practical scenarios.

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 hope that these will be addressed in the future.

Previous works

Linear cryptanalysis for block ciphers was introduced by Matsui in [29] using a single linear approximation. Subsequently, 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 the attacks [30, 25] 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 [33].

Differential cryptanalysis was proposed by Biham and Shamir [6] where a single differential was used. In a later work [7], the authors improved the attack by considering together several differentials having the same output difference. Kundsen [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 in [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, 15, 27]. These works also require approximations and the message of the present work should also be applicable to these works.

Summary 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 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 [33] and [20] using the hypothesis testing based framework. Section 7 does the same for the results on single and multiple differential cryptanalysis appearing in [33] and [11]. 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 and 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-Esséen theorem which is a standard result in probability theory [5, 17, 18]. The form of this theorem that is required for our purpose is stated as follows [19].

Theorem 1. *Suppose $V_1, \ldots, V_\lambda$ are independent and identically distribution random variables with $E[V_i] = 0$, $E[V_i^2] = \sigma^2$ and $E[|V_i|^3] = \rho < \infty$. Let $D_\lambda = (V_1 + \cdots + V_\lambda)/(\sqrt{\lambda}\sigma)$ and $\Phi()$ 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*

$$|\Pr[D_\lambda \leq x] - \Phi(x)| \leq \frac{C\rho}{\sigma^3\sqrt{\lambda}}. \tag{1}$$

The best known upper and lower bounds on C are $C < 0.4748$ [34] and $C \geq 0.40973$ [18] respectively. This result 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 1. *Let $Y_1, \ldots, Y_\lambda$ be a sequence of independent* Bernoulli(p) *random variables, with $p \in (0, 1)$ and let $c = 2(p - 1/2)$. Then for every real x and positive integer λ,*

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

where

$$\Lambda_p \quad = \quad \frac{C \times (p^2 + (1-p)^2)}{(p(1-p))^{1/2}} = \frac{C \times (1 + c^2)}{((1 - c^2))^{1/2}}. \tag{3}$$

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

$$\lambda \geq 2^{2\varepsilon}\Lambda_p^2 = 2^{2\varepsilon}C^2 \times \frac{(p^2 + (1-p)^2)^2}{p(1-p)} = 2^{2\varepsilon}C^2 \times \frac{(1 + c^2)^2}{1 - c^2}. \tag{4}$$

Note: In (1) and (2) the difference between the two probabilities is given by a 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) is at most $2^{-\varepsilon}$. This is ensured by requiring the upper bound in (2) to be at most $2^{-\varepsilon}$. Depending on the value of p, this gives rise to two types of conditions.

1. **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 side of (2) is $\approx 2^{-\varepsilon}$. In this case, the value of λ for which the approximation is good is independent of c.

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

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 [35] 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. *Let $T_1, \ldots, T_\lambda$ 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 $T_{(1)} \leq T_{(2)} \leq \ldots \leq T_{(\lambda)}$ be the order statistics of $T_1, \ldots, T_\lambda$. 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*

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

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

$$\left| \Pr[W_\lambda \leq z] - \Phi\left(\frac{z f(\xi_q)}{\sigma_\lambda} + \frac{(\zeta_\lambda - \lambda^{-1/2}(r(\lambda) - \lambda q))}{\sigma_\lambda} \right) \right| \quad \leq \quad \frac{\Lambda_{q_\lambda}}{\lambda^{1/2}}. \tag{5}$$

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

$$\Lambda_{q_\lambda} \quad = \quad C \times \frac{(1-q_\lambda)^2 + q_\lambda^2}{(q_\lambda(1-q_\lambda))^{1/2}} \tag{6}$$

and $\Phi()$ is the standard normal distribution function.

Consequently, as $\lambda \to \infty$ 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

$$\Pr[Z_\lambda \leq z] \quad = \quad \Pr\left[T_{(r(\lambda))} \leq \xi_q + \frac{z}{\lambda^{\frac{1}{2}} f(\xi_q)} \right]. \tag{7}$$

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

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

This shows that the U_i's 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 $\{1, \ldots, n\}$ such that $T_{(i)} = T_{\pi(i)}$ then $U_{(i)} = U_{\pi(i)}$, i.e., the U_i's preserve the order of the T_i's 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 (7) we obtain:

$$\Pr[Z_\lambda \leq z] \quad = \quad \Pr[U_{(r(\lambda))} \leq q_\lambda]. \tag{8}$$

Define binary valued random variables $Y_1, \ldots, Y_\lambda$ such that $Y_i = 1$ if $U_i \leq q_\lambda$ and is 0 otherwise; further define $A_\lambda = Y_1 + \cdots + Y_\lambda = \#\{i : U_i \leq q_\lambda\}$. The variables Y_i's are independent $\mathsf{Bernoulli}(q_\lambda)$ distributed. Let $\sigma_\lambda^2 = E[(Y_i - q_\lambda)^2]$ and $\rho_\lambda = E[|Y_i - q_\lambda|^3]$.

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

$$\Pr[Z_\lambda \leq z] \quad = \quad \Pr[U_{(r(\lambda))} \leq q_\lambda] = \Pr[A_\lambda \geq r(\lambda)]. \tag{9}$$

We now propose to use the Berry-Esséen theorem as given in Corollary 1. Define

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

From Corollary 1, for every real x and positive integer λ,

$$\frac{C \times \rho_\lambda}{\sigma_\lambda^3 \lambda^{1/2}} \geq |\Pr[D_\lambda < x] - \Phi(x)| = |\Pr[D_\lambda \geq x] - \Phi(-x)|\,. \tag{10}$$

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

$$\Pr[Z_\lambda \leq z] = \Pr[A_\lambda \geq r(\lambda)] = \Pr\left[D_\lambda \geq \lambda^{-1/2} \frac{r(\lambda) - \lambda q_\lambda}{\sigma_\lambda}\right]. \tag{11}$$

We now take a more careful look at q_λ. Define

$$\zeta_\lambda \quad = \quad \frac{z}{f\left(\xi_q\right)} \left(\frac{F\left(\xi_q + \frac{z}{\lambda^{\frac{1}{2}} f(\xi_q)}\right) - q}{\frac{z}{\lambda^{\frac{1}{2}} f(\xi_q)}} - f\left(\xi_q\right) \right). \tag{12}$$

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

$$\lim_{\lambda \to \infty} \zeta_\lambda = \lim_{\lambda \to \infty} \frac{z}{f(\xi_q)} \left(\frac{F\left(\xi_q + \Delta x\right) - q}{\Delta x} - f(\xi_q) \right) = \frac{z}{f(\xi_q)} \left(\lim_{\Delta x \to 0} \frac{F\left(\xi_q + \Delta x\right) - q}{\Delta x} - f(\xi_q) \right) = 0. \tag{13}$$

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 (12) we can write

$$q_\lambda = F\left(\xi_q + z/(\lambda^{1/2} f(\xi_q))\right) = q + \frac{1}{\sqrt{\lambda}}(z + \zeta_\lambda). \tag{14}$$

So, we have q_λ tends to q as λ tends to infinity. Using the expression for q_λ given by (14) in (11) we obtain

$$\Pr[Z_\lambda \leq z] \quad = \quad \Pr[D_\lambda \geq (\lambda^{-1/2}(r(\lambda) - \lambda q) - \zeta_\lambda)/\sigma_\lambda - z/\sigma_\lambda]. \tag{15}$$

Setting $x = \lambda^{-1/2}(r(\lambda) - \lambda q)/\sigma_\lambda - \zeta_\lambda/\sigma_\lambda - z/\sigma_\lambda$ and substituting in (10) we obtain

$$\left|\Pr[Z_\lambda \leq z] - \Phi\left(z/\sigma_\lambda + (\zeta_\lambda - \lambda^{-1/2}(r(\lambda) - \lambda q))/\sigma_\lambda\right)\right| \leq \frac{C \times \rho_\lambda}{\sigma_\lambda^3 \lambda^{1/2}}. \tag{16}$$

Note that $W_\lambda = Z_\lambda / f(\xi_q)$. Then from (16) we obtain

$$\left|\Pr[W_\lambda \leq z] - \Phi\left(z f(\xi_q)/\sigma_\lambda + (\zeta_\lambda - \lambda^{-1/2}(r(\lambda) - \lambda q))/\sigma_\lambda\right)\right| \leq \frac{C \times \rho_\lambda}{\sigma_\lambda^3 \lambda^{1/2}}. \tag{17}$$

This shows the desired bound.

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

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 $E : \{0,1\}^k \times \{0,1\}^n \to \{0,1\}^n$ such that for each $K \in \{0,1\}^k$, the function $E_K(\cdot) \stackrel{\Delta}{=} E(K,\cdot)$ 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 the round keys as $k^{(0)}, k^{(1)}, \ldots$, and the round functions as $R^{(0)}_{k^{(0)}}, R^{(1)}_{k^{(1)}}, \ldots$. The round functions are bijections of $\{0,1\}^n$. Further, denote by $K^{(i)}$ the concatenation of the first i round keys and $E^{(i)}_{K^{(i)}}$ the composition of the first i round functions, i.e.,

$$E^{(i)}_{K^{(i)}} = R^{(i)}_{k^{(i)}} \circ \cdots \circ R^{(0)}_{k^{(0)}} = R^{(i)}_{k^{(i)}} \circ E^{(i-1)}_{K^{(i-1)}}.$$

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, $B = E^{(r)}_{K^{(r)}}(P)$ and $C = R^{(r)}_{k^{(r)}}(B)$.

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 one 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

$$M \;=\; 2^m - 1 \tag{18}$$

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), $i = 1, \ldots, N$ which have been generated using the correct key and are available. The plaintexts $P_1, \ldots, P_N$ are supposed to be independent and uniformly distributed over the plaintext space. 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 depend on whether κ is correct or incorrect. It is also assumed that the distributions of T_κ for incorrect κ are identical. 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 $i = 1, \ldots, \ell$. The subscript P denotes plaintext mask; the subscript B denotes 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

$$L_i = \langle \Gamma_P^{(i)}, P \rangle \oplus \langle \Gamma_B^{(i)}, B \rangle; \text{ for } i = 1, \ldots, \ell. \tag{19}$$

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 $z = (z_1, \ldots, z_\ell)$ the collection of the ℓ bits arising in this manner. The key masks $\Gamma_K^{(1)}, \ldots, \Gamma_K^{(\ell)}$ 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:

$$X \quad = \quad (L_1, \ldots, L_\ell). \tag{20}$$

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

A single linear approximation is of the form

$$L_i = \langle \Gamma_K^{(i)}, K^{(r)} \rangle = z_i. \tag{21}$$

Note that we are not assuming any randomness over the key $K^{(r)}$ and so the bits z_i's 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 (21) holds with some probability over the uniform random choice of P. The random variables $L_1, \ldots, L_\ell$ are not necessarily independent. The joint distribution of these variables is given as follows: For $z = (z_1, \ldots, z_\ell)$, and $\eta = (\eta_1, \ldots, \eta_\ell) \in \{0,1\}^\ell$, define

$$p_z(\eta) \quad = \quad \Pr[L_1 = \eta_1 \oplus z_1, \ldots, L_\ell = \eta_\ell \oplus z_\ell] = \frac{1}{2^\ell} + \epsilon_\eta(z) \tag{22}$$

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

For each choice of z, we obtain the distribution $\tilde{p}_z \stackrel{\Delta}{=} (p_z(0), \ldots, p_z(2^\ell - 1))$, where the integers $\{0, \ldots, 2^\ell - 1\}$ are identified with the set $\{0,1\}^\ell$. Let $\tilde{p}$ be the probability distribution $\tilde{p} \stackrel{\Delta}{=} \tilde{p}_{0^\ell}$ so that

$$\tilde{p}(\eta) \;=\; 1/2^\ell + \epsilon_\eta \text{ for } \eta \in \{0,1\}^\ell. \tag{23}$$

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

$$C(\tilde{p}) \;=\; 2^\ell \times \sum_{\eta \in \{0,1\}^\ell} \epsilon_\eta^2. \tag{24}$$

A potential conflict of notation: It is customary in the cryptanalysis literature to denote capacity by $C(\tilde{p})$. Similarly, it is customary in the literature on mathematical statistics to denote the constant in the Berry-Esséen theorem by C. This creates a potential conflict of notation. Denoting either of these quantities by a different symbol will be inconsistent with the respective literature. So, we decided to not change the conventions. Since capacity $C(\tilde{p})$ will always be written as a function applied to a probability distribution $\tilde{p}$, hopefully, it will be possible to differentiate it from just the symbol C used to denote the constant in the Berry-Esséen theorem.

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

$$L_{\kappa,j,i} \;=\; \langle \Gamma_P^{(i)}, P_j \rangle \oplus \langle \Gamma_B^{(i)}, B_{\kappa,j} \rangle; \tag{25}$$

$$X_{\kappa,j} \;=\; (L_{\kappa,j,1}, \ldots, L_{\kappa,j,\ell}). \tag{26}$$

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

$$Q_{\kappa,\eta} \;=\; \#\{j : X_{\kappa,j} = \eta\}, \text{ for } \eta = 0, \ldots, 2^\ell - 1. \tag{27}$$

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

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 [33] 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 $T_0, \ldots, T_{2^m-1}$ and assumes T_0 to be the test statistics corresponding to the correct choice of κ. Let $T_{(1)}, \ldots, T_{(M)}$ be the order statistics of $T_1, \ldots, T_M$. Recall $M = 2^m - 1$.

Advantage of an attack: Selçuk defines an attack to have an a-bit advantage if $T_0 > T_{(2^m(1-2^{-a}))}$, 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 $T_0 > T_{(2^m q)}$ where $q = 1 - 2^{-a}$.

Selçuk's key idea was to use Theorem 2 to approximate the distribution of $T_{(2^m q)}$. For this, the value of λ in Theorem 2 is set to M and 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 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 $T_{(2^m q)}$ approximately follows $\mathcal{N}(\xi_q, q(1-q)/(\lambda f^2(\xi_q)))$. Using $\lambda = M - 1 = 2^m - 1$, the random variable $T_{(2^m q)}$ 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 $T_0 > T_{(2^m q)}$. So, the probability of success is the probability that $T_0 - T_{(2^m q)} > 0$. To compute this probability, the distribution of $T_0 - T_{(2^m q)}$ is required. The result on order statistics shows that $T_{(2^m q)}$ approximately follows a normal distribution. If it also turns out that T_0 follows a normal distribution, then $T_0 - T_{(2^m q)}$ 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 normal, then in general it is not clear how to obtain the distribution for $T_0 - T_{(2^m q)}$. 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.

4.2 Convergence and Tightness Issues

Consideration of convergence and tightness issues give rise to two constraints.

Constraint 1. m **must be large:** In the proof of Theorem 2, for the convergence in distribution to $\mathcal{N}(0, q(1-q)/f^2(\xi_q))$ it is required that $q_\lambda \to q$ as $\lambda \to \infty$. This raises the question of how large should λ 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 $q_\lambda \to q$ as $\lambda \to \infty$. 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. Under the approximation $q_\lambda \approx q$, denote the quantity Λ_{q_λ} given by (6) 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 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

$$2^{m-a} \;\geq\; 2^{2\varepsilon} \times C^2 \times \frac{(2^{-2a} + (1-2^{-a})^2)^2}{1-2^{-a}}. \tag{28}$$

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 $(2^{-2a} + (1-2^a)^2)^2/(1-2^{-a})$ is a

small constant which gets close to 1 as a gets close to m. So, from (28) it follows that for the error in Theorem 2 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.

4.3 Selçuk's Method of Linear Cryptanalysis: A Critique

Let us briefly consider Selçuk's method of linear cryptanalysis. 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 $W_\kappa = (L_{\kappa,1} + \cdots + L_{\kappa,N})/N - 1/2$. 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 $T_0 > T_{(2^m q)}$ where $q = 1 - 2^{-a}$. The distribution of $T_{(2^m q)}$ 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 $\mathsf{Bernoulli}(1-p)$. Using Corollary 1, 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 $T_0 > T_{(2^m q)}$. With T_0 following a folded normal distribution and $T_{(2^m q)}$ approximately following a normal distribution, it becomes difficult to obtain the distribution of $T_0 - T_{(2^m q)}$.

Selçuk instead considers the event success to be $W_0 > T_{(2^m q)}$. 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.

$$\begin{aligned}
\Pr[T_0 > T_{(2^m q)}] &= \Pr[|W_0| > T_{(2^m q)}] \\
&= \Pr[W_0 > T_{(2^m q)} | W_0 > 0] \times \Pr[W_0 \geq 0] + \Pr[-W_0 > T_{(2^m q)} | W_0 < 0] \times \Pr[W_0 < 0] \\
&\approx \Pr[W_0 > T_{(2^m q)} | W_0 > 0] \times \Pr[W_0 \geq 0] \\
&= \Pr[W_0 > T_{(2^m q)}]. \qquad (29)
\end{aligned}$$

Without the assumption $\Pr[W_0 < 0] \approx 0$, we see no justification for considering success to be the event $W_0 > T_{(2^m q)}$. 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 $c = 2(p - 1/2)$ and since we are assuming $p > 1/2$, it follows that $c > 0$. Then,

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

So, N has to be sufficiently large to ensure that the probability of $W_0 < 0$ is negligible. For example, $\Phi(-5) \approx 2.86651571879 \times 10^{-7}$. Considering 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 $N > 25(1-c^2)/c^2$.

Note that in the context of linear cryptanalysis, c is likely to be a small value. So, the requirement that $N > 25(1-c^2)/c^2$ puts a non-trivial requirement on the data complexity N. In other words, the condition that

T_0 is almost surely non-negative puts a lower bound on the required number of plaintext-ciphertext pairs. It may turn out that this lower bound on N is higher than the lower bound on N obtained 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 [33], 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 missed in [33].

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:

$$\left.\begin{array}{l} H_0 : \kappa \text{ is correct; versus } H_1 : \kappa \text{ is incorrect.} \\ \text{Decision rule: Reject } H_0 \text{ if } T_\kappa \leq t. \end{array}\right\} \tag{30}$$

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.

$$\left.\begin{array}{rcl} \Pr[\text{Type-I error}] & = & \Pr[\kappa \text{ is rejected } | H_0 \text{ holds }] = \Pr[T \leq t | H_0 \text{ holds }] \leq \alpha; \\ \Pr[\text{Type-II error}] & = & \Pr[\kappa \text{ is accepted } | H_1 \text{ holds }] = \Pr[T > t | H_1 \text{ holds }] \leq \beta; \\ \Pr[\mathsf{succ}] & = & 1 - \Pr[\text{Type-I error}] \geq 1 - \alpha = P_S. \end{array}\right\} \tag{31}$$

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 on the probabilities of Type-I and Type-II errors. If α is a 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 N. Eliminating N between these two expressions provide the expression for the threshold parameter t in terms of P_S and β. On the other hand, eliminating t between these two expressions provide 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

$$\beta \quad = \quad 2^{-a}. \tag{32}$$

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 [33] and that of multiple linear cryptanalysis was considered in [20]. The analysis of Section 4 puts a question mark on the framework of 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 [33] 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 $j \in \{1, \ldots, N\}$, under the null hypothesis H_0 (i.e., for the correct choice of κ), $L_{\kappa,j}$ follows $\mathsf{Bernoulli}(1-p)$ for each j and under the alternate hypothesis H_1 (i.e., for an incorrect choice of κ), $L_{\kappa,j}$ follows $\mathsf{Bernoulli}(1/2)$.

Let $\mu_0 = p$, $\sigma_0^2 = p(1-p)$ and $\mu_1 = 1/2$, $\sigma_1^2 = 1/4$ denote the means and the variances 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:

$$T_\kappa \quad = \quad |Z_\kappa| \text{ where } Z_\kappa = \frac{(L_{\kappa,1} + \cdots + L_{\kappa,N})/N - \mu_1}{\sigma_1/\sqrt{N}}. \tag{33}$$

The hypothesis testing framework and the decision rule mentioned in (30) is applied to T_κ.

From Corollary 1, 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 $2^{-\varepsilon}$, then it is sufficient to have $N > 2^{2\varepsilon}C^2$.

We write

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

Again from Corollary 1, 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

$$N \geq 2^{2\varepsilon} C^2 (1+c^2)^2/(1-c^2) \tag{34}$$

the error in approximation will be at most $2^{-\varepsilon}$. If p is close to $1/2$ (which is often the case) and so c is close to 0, then $N \approx 2^{2\varepsilon}$ ensures that the approximation error is around $2^{-\varepsilon}$. More generally, one has to consider the bound given by (34).

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

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

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

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

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

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

We set

$$\alpha = \Phi\left(\frac{\sigma_1 t}{\sigma_0} - \frac{|\mu_0-\mu_1|\sqrt{N}}{\sigma_0}\right) \text{ and } P_S = 1-\alpha. \tag{35}$$

So, for any $\mu_0 \neq \mu_1$, Pr[Type-I error] $\leq \alpha$ the probability of success is at least P_S. Using (35)

$$\sigma_1 t = \sigma_0 \Phi^{-1}(1-P_S) + |\mu_0-\mu_1|\sqrt{N} = -\sigma_0 \Phi^{-1}(P_S) + |\mu_0-\mu_1|\sqrt{N}. \tag{36}$$

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

$$\begin{aligned}\beta = \Pr[\text{Type-II error}] &= \Pr\left[|T_\kappa| > t \,|H_1 \text{ holds}\right] = \Pr\left[T_\kappa < -t \,|H_1 \text{ holds}\right] + \Pr\left[T_\kappa > t \,|H_1 \text{ holds}\right] \\ &= \Phi(-t) + 1 - \Phi(t) = 2\left(1-\Phi(t)\right).\end{aligned} \tag{37}$$

Eliminating t using (36) and (37); substituting the values of σ_0, σ_1 and $(\mu_0-\mu_1)^2$ in terms of c; and using (32) we obtain

$$N = \frac{\left\{\sigma_1 \Phi^{-1}(1-\beta/2) + \sigma_0 \Phi^{-1}(P_S)\right\}^2}{(\mu_0-\mu_1)^2} = \frac{\left\{\Phi^{-1}\left(1-2^{-a-1}\right) + \sqrt{1-c^2}\Phi^{-1}(P_S)\right\}^2}{c^2}. \tag{38}$$

In [33], Selçuk makes the assumption $1-c^2 \approx 1$. Using this assumption, the expression for N given by (38) becomes exactly the same as that obtained in [33].

It will usually be the case that for a reasonable value of ε, the expression for N given by (38) is larger than the bound given by (34). 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 (38) and (34).

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 (27) to be $Q_{\kappa,\eta} = \#\{j \in \{1,\ldots,N\} : X_{\kappa,j} = \eta\}$, i.e., $Q_{\kappa,\eta}$ is the number of times η appears among the random variables $X_{\kappa,1},\ldots,X_{\kappa,N}$.

6.2.1 Data Complexity for the χ^2 Test Statistics

The test statistics used by Hermelin et al [20] was

$$T_\kappa \;=\; 2^\ell N \sum_{\eta=0}^{2^\ell-1} \left(Q_{\kappa,\eta}/N - 2^{-\ell}\right)^2. \tag{39}$$

We use T_κ in the hypothesis testing framework given in (30). The distribution of T_κ is approximated in [20] as follows. This is justified by a reference to [16]. Later we consider this issue in more details.

$$\left.\begin{array}{lcll} T_\kappa & \sim & \chi^2_{2^\ell-1}(NC(\tilde{p})) & \text{under } H_0; \\ T_\kappa & \sim & \chi^2_{2^\ell-1} & \text{under } H_1. \end{array}\right\} \tag{40}$$

It is difficult to work with a non-central chi-squared distribution. So, [20] approximates $\chi^2_{2^\ell-1}\left(NC\left(\tilde{p}\right)\right)$ for large values of $2^\ell - 1 + NC\left(\tilde{p}\right)$ by normal $\mathcal{N}\left(2^\ell - 1 + NC\left(\tilde{p}\right), 2\left(2^\ell - 1 + 2NC\left(\tilde{p}\right)\right)\right)$ 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 $\chi^2_{2^\ell-1}$ distribution. Then

$$\left.\begin{array}{lcl} \alpha & = & \Pr[\text{Type-I error}] = \Pr\left[T_\kappa \leq t \,|H_0 \text{ holds}\right] = \Pr\left[T_\kappa \leq t\right] = \Phi\left(\frac{t-\left(2^\ell-1+NC(\tilde{p})\right)}{\sqrt{2\left(2^\ell-1+2NC(\tilde{p})\right)}}\right) = 1 - P_S. \\ \beta & = & \Pr[\text{Type-II error}] = \Pr\left[T_\kappa > t \,|H_1 \text{ holds}\right] = 1 - \Psi_{2^\ell-1}\left(t\right). \end{array}\right\} \tag{41}$$

So, $t = 2^\ell - 1 + NC\left(\tilde{p}\right) + \sqrt{2\left(2^\ell - 1 + 2NC\left(\tilde{p}\right)\right)}\Phi^{-1}\left(1 - P_S\right) = \Psi^{-1}_{2^\ell-1}(1-\beta)$. Simplifying leads to the following relation which is a quadratic equation in $(NC(\tilde{p}))$.

$$\left(NC\left(\tilde{p}\right)\right)^2 - 2(NC\left(\tilde{p}\right))\left\{\gamma - 2^\ell - 1 + 2\delta^2\right\} + \left\{\gamma - 2^\ell - 1\right\}^2 - 2(2^\ell - 1)\delta^2 = 0, \tag{42}$$

where $\gamma = \Psi^{-1}_{2^\ell-1}(1-\beta)$ 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 (32) to substitute 2^{-a} for β we obtain:

$$N = \frac{1}{C(\tilde{p})}\left[2\left\{\Phi^{-1}(P_S)\right\}^2 + \Psi^{-1}_{2^\ell-1}\left(1-2^{-a}\right) - (2^\ell - 1) + \left|\Phi^{-1}(P_S)\right|\sqrt{2\Psi^{-1}_{2^\ell-1}\left(1-2^{-a}\right) - 4\left\{\Phi^{-1}(P_S)\right\}^2}\right]. \tag{43}$$

Approximating $\chi^2_{2^\ell-1}$ by $\mathcal{N}(2^\ell - 1, 2(2^\ell - 1))$: The expression for N given by (43) 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 $\chi^2_{2^\ell-1}$ by $\mathcal{N}(2^\ell - 1, 2(2^\ell - 1))$ as has been done in [20]. The expression for β now becomes $\beta = \Pr[\text{Type-II error}] = 1 - \Phi\left((t - (2^\ell - 1))/\sqrt{2(2^\ell - 1)}\right)$. Combining this with the expression for P_S given by (41) we obtain the following quadratic in $(NC(\tilde{p}))$.

$$\left(NC(\tilde{p})\right)^2 - 2N(C(\tilde{p})\left(\sqrt{2(2^\ell - 1)}\gamma + 2\delta^2\right) + 2(2^\ell - 1)\left(\gamma^2 - \delta^2\right) = 0. \tag{44}$$

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 (32) to substitute 2^{-a} for β we obtain:

$$N = \frac{1}{C(\tilde{p})}\left[\sqrt{2(2^\ell-1)}\Phi^{-1}\left(1-2^{-a}\right) + 2\left(\Phi^{-1}\left(P_S\right)\right)^2 + \right.$$
$$\left.\Phi^{-1}\left(P_S\right)\sqrt{4\left(\Phi^{-1}\left(P_S\right)\right)^2 + 4\sqrt{2(2^\ell-1)}\Phi^{-1}\left(1-2^{-a}\right) + 2(2^\ell-1)}\right]. \quad (45)$$

It was mentioned in [20] that 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 $P_S = \Phi\left((\mu_R - \mu_q)/\left((\sigma_R^2+\sigma_q^2)^{1/2}\right)\right)$. From Section 5.1 of [20], $\mu_R = (2^\ell-1)+NC\left(\tilde{p}\right), \sigma_R^2 = 2\left((2^\ell-1)+2NC(\tilde{p})\right), \mu_q = \sqrt{2(2^\ell-1)}\Phi^{-1}\left(1-2^{-a}\right)+(2^\ell-1)$ and $\sigma_q^2 = 2^{-(l+a)}\sigma_w^2/\phi^2\left(\Phi^{-1}\left(1-2^{-a}\right)\right) \ll \sigma_R^2$. 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 (44). Consequently, the expression for the data complexity N in this case is also given by (45). In [20], the authors applied some further approximations to (45) to arrive at a simplified expression for the data complexity. It was noted in [20] 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's 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 $(L_1 \oplus z_1, \ldots, L_\ell \oplus z_\ell) = \eta$. For a choice κ of the last round key bits and z of the inner key bits we make the following definitions:

$$\begin{aligned} X_{\kappa,z,j} &= (L_{\kappa,j,1} \oplus z_1, \ldots, L_{\kappa,j,\ell} \oplus z_\ell); \\ Q_{\kappa,z,\eta} &= \#\{j : X_{\kappa,z,j} = \eta\}, \text{ for } \eta = 0, \ldots, 2^\ell - 1. \end{aligned}$$

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 and the alternate hypothesis H_0 and H_1 respectively remain unchanged. The test statistic and the decision rule change. We denote the test statistics by $T_{\kappa,z}$ which is defined as follows.

$$T_{\kappa,z} = \sum_{\eta \in \{0,1\}^\ell} Q_{\kappa,z,\eta} \log \frac{\tilde{p}_z(\eta)}{2^{-\ell}}. \quad (46)$$

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

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

Later, we consider the error in the approximation.

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

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

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

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

We define $\alpha = \Pr[T_{\kappa,z^*} \leq t | H_0 \text{ holds}]$ 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^*} follows $\mathcal{N}(N\mu_0, N\sigma_0^2)$. From this we obtain

$$\begin{aligned} & 1 - P_S = \alpha = \Pr[T_{\kappa,z^*} \leq t | H_0 \text{ holds}] = \Phi\left(\frac{t - N\mu_0}{\sqrt{N}\sigma_0}\right), \\ \Rightarrow \quad & t = N\mu_0 + \sqrt{N}\sigma_0 \Phi^{-1}(1 - P_S). \end{aligned} \tag{47}$$

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

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

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 (47); substituting the values of μ_0, μ_1, σ_0 and σ_1; and then using (32) to substitute 2^{-a} for β provides the following expression for the data complexity N.

$$N = \frac{\left\{\Phi^{-1}\left(1 - 2^{-\ell-a}\right) + \Phi^{-1}(P_S)\right\}^2}{C(\tilde{p})}. \tag{48}$$

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

$$\frac{\left\{\Phi^{-1}(P_{12}) + \Phi^{-1}\left(\sqrt[2^\ell]{1 - 2^{-a}}\right)\right\}^2}{C(\tilde{p})}. \tag{49}$$

Here P_{12} is a lower bound on the success probability and can be taken to be P_S. In [20], the authors had used the approximation $\sqrt[2^\ell]{1 - 2^{-a}} \approx 1 - 2^{-a-\ell}$. Using this approximation in (49) gives the expression for N to be exactly the same as that given in (48).

7 Differential Cryptanalysis

Expressions for the data complexities of differential cryptanalysis were derived in [33] 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 $\delta_0, \delta_1, \ldots, \delta_r$ 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 $B^{(0)} = P, B^{(1)}, \ldots, B^{(r)}$ be the inputs to round numbers $0, \ldots, r$ respectively, i.e., $B^{(i+1)} = R^{(i)}_{k^{(i)}}(B^{(i)})$ and suppose that $B^{(0)\prime} = P', B^{(1)\prime}, \ldots, B^{(r)\prime}$ be the inputs to round numbers $0, \ldots, r$ respectively corresponding to the plaintext P'. Consider the event $A = \wedge_{i=0}^{r}(B^{(i)} \oplus B^{(i)\prime} = \delta_i)$ 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 $\delta_0, \delta_1, \ldots, \delta_r$; 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 ciphertext C and C' corresponding to the plaintexts P and P' respectively. Guessing a target sub-key allows computing $B^{(r)}$ and $B^{(r)\prime}$ respectively from C and C'. The earlier $B^{(i)}$'s and $B^{(i)\prime}$'s cannot be obtained. Define the event D to be $B^{(r)} \oplus B^{(r)\prime} = \delta_r$. The event A implies the event D. The event D may also occur even when $\overline{A}$ 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 $\Pr[D|\overline{A}] = p'$ and let $p_0 = p + (1-p)p'$. Then for the correct choice κ of the target sub-key

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

Since δ_0 is not the zero string, $P \neq P'$ and since each round function is a bijection $B^{(i)} \neq B^{(i)\prime}$ for $i = 1, \ldots, r$. If the choice κ of the target sub-key is incorrect, then it is reasonable to assume that $B^{(r)}$ and $B^{(r)\prime}$ 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 $\Pr[D] = 1/(2^m - 1)$. Let $p_w = 1/(2^m - 1)$. In general $p_0 > p_w$.

For the actual attack, there will be N matched pairs of plaintexts $(P_1, P_1'), \ldots, (P_N, P_N')$ with $P_j \oplus P_j' = \delta_0$ and their corresponding ciphertexts $(C_1, C_1'), \ldots, (C_N, C_N')$. For a choice κ of the target sub-key, we obtain $(B^{(r)}_{\kappa,1}, B^{(r)\prime}_{\kappa,1}), \ldots, (B^{(r)}_{\kappa,N}, B^{(r)\prime}_{\kappa,N})$ by inverting $(C_1, C_1'), \ldots, (C_N, C_N')$ respectively. So, it is possible to determine whether the condition $B^{(r)}_{\kappa,j} \oplus B^{(r)\prime}_{\kappa,j} = \delta_r$ holds for each $j = 1, \ldots, N$.

For a choice κ of the target sub-key, and $j = 1, \ldots, N$, define random variables $W_{\kappa,j} = 1$ if $B^{(r)}_{\kappa,j} \oplus B^{(r)}_{\kappa,j} = \delta_r$; and $W_{\kappa,j} = 0$ otherwise. If κ is the correct choice, the $\Pr[W_{\kappa,j} = 1] = p_0$ and if κ is an incorrect choice, then $\Pr[W_{\kappa,j} = 1] = p_w$ for all j.

The hypothesis testing framework in (30) is to be applied. Define the test statistics

$$T_\kappa = \frac{(W_{\kappa,1} + \cdots + W_{\kappa,N}) - Np_w}{\sqrt{Np_w(1-p_w)}}. \tag{50}$$

From Corollary 1, under H_1, T_κ is approximated by the standard normal distribution for sufficiently large N. From the error bound, we have that if

$$N \geq 2^{2\varepsilon}C^2\frac{(p_w^2 + (1-p_w)^2)^2}{p_w(1-p_w)} \tag{51}$$

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

$$T_\kappa = \frac{(W_{\kappa,1} + \cdots + W_{\kappa,N}) - Np_w}{\sqrt{Np_w(1-p_w)}} = Y_\kappa \times \frac{\sqrt{Np_0(1-p_0)}}{\sqrt{Np_w(1-p_w)}} + \frac{Np_0 - Np_w}{\sqrt{Np_w(1-p_w)}}.$$

where

$$Y_\kappa = \frac{(W_{\kappa,1} + \cdots + W_{\kappa,N}) - Np_0}{\sqrt{Np_0(1-p_0)}}$$

Again from Corollary 1, under H_0, the distribution of Y_κ is approximated by the standard normal distribution. Further, if

$$N \geq 2^{2\varepsilon} C^2 \frac{(p_0^2 + (1-p_0)^2)^2}{p_0(1-p_0)} \tag{52}$$

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

$$\begin{aligned}
\alpha = \Pr[\text{Type-I Error}] &= \Pr\left[T_\kappa \leq t \,|H_0 \text{ holds}\right] \\
&= \Pr\left[Y_\kappa \leq \left(\frac{\sqrt{Np_w\,(1-p_w)}}{\sqrt{Np_0\,(1-p_0)}}\right)\left(t - \frac{Np_0 - Np_w}{\sqrt{Np_w\,(1-p_w)}}\right)\right] \\
&= \Pr\left[Y_\kappa \leq \frac{t\sqrt{Np_w\,(1-p_w)}}{\sqrt{Np_0\,(1-p_0)}} - \frac{Np_0 - Np_w}{\sqrt{Np_0\,(1-p_0)}}\right] \\
&= \Phi\left(\frac{t\sqrt{Np_w\,(1-p_w)}}{\sqrt{Np_0\,(1-p_0)}} - \frac{Np_0 - Np_w}{\sqrt{Np_0\,(1-p_0)}}\right) = 1 - P_S.
\end{aligned} \tag{53}$$

Type-II error is given by,

$$\beta = \Pr[\text{Type-II Error}] = \Pr\left[T_\kappa > t \,|H_1 \text{ holds}\right] = 1 - \Phi(t). \tag{54}$$

which implies $t = \Phi^{-1}(1-\beta)$. Using this value of t in (53) and using (32) 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 [33] 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

$$N = \left\{\frac{\sqrt{p_w}\Phi^{-1}(1-2^{-a}) + \sqrt{p_0}\Phi^{-1}(P_S)}{p_0 - p_w}\right\}^2. \tag{55}$$

This is identical to the expression for data complexity obtained by Selçuk in [33].

The expression for N given by (55) has to be compared with the lower bounds for N given by (51) and (52). Usually, it will be the case that $p_0 > p_w$ and so it is sufficient to only consider the lower bound given by (51). The right hand side of (55) is inversely proportional to $(p_0 - p_w)^2$ whereas the right hand side of (51) 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 (55) to be greater than the one given by (51). 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 [33] had 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 (55) 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 ν sequences of n-bit strings $\delta_j^{(1)}, \ldots, \delta_j^{(\nu)}$ for $j = 1, \ldots, r$. 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'), $j = 1, \ldots, N$ 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, $j = 1, \ldots, N$, 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, $i = 1, \ldots, \nu$, to be the probability that $B_{\kappa,j}^{(r)} \oplus B_{\kappa,j}^{(r)\prime}$ equals $\delta_r^{(i)}$. This probability is independent of both j and κ. For an incorrect choice κ, the probability that $B_{\kappa,j}^{(r)} \oplus B_{\kappa,j}^{(r)\prime}$ equals $\delta_r^{(i)}$ is $1/(2^m - 1)$. Define $p_0 = 1 - \sum_{i=1}^{\nu} p_i$ and $\theta_0 = 1 - \nu/(2^m - 1)$. For the correct choice of κ, p_0 represents the probability that none of the events $B_{\kappa,j}^{(r)} \oplus B_{\kappa,j}^{(r)\prime} = \delta_r^{(i)}$ occur; similarly, for an incorrect choice of κ, θ_0 represents the probability that none of the events $B_{\kappa,j}^{(r)} \oplus B_{\kappa,j}^{(r)\prime} = \delta_r^{(i)}$ occur. So, $(p_0, p_1, \ldots, p_\nu)$ and $(\theta_0, 1/(2^m - 1), \ldots, 1/(2^m - 1)$ are both probability distributions corresponding respectively to the correct and the incorrect choices of the target sub-key. For convenience of notation we assume $\theta_i = 1/(2^m - 1)$ for $i = 1, \ldots, \nu$ and $\tilde{p} = (p_0, p_1, \ldots, p_\nu)$.

7.3 The LLR Test Statistic

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

$$\begin{aligned} X_{\kappa,i} &= \#\left\{j : B_{\kappa,j}^{(r)} \oplus B_{\kappa,j}^{(r)} = \delta_r^{(i)}\right\}; \\ T_\kappa &= \sum_{i=0}^{\nu} X_{\kappa,i} \log\left(\frac{p_i}{\theta_i}\right). \end{aligned}$$

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

$$\mu = \begin{cases} \mu_0 \approx \sum_{i=0}^{\nu} p_i \log(p_i/\theta_i) & \text{if } H_0 \text{ holds;} \\ \mu_1 \approx \sum_{i=0}^{\nu} \theta_i \log(\theta_i/p_i) & \text{if } H_1 \text{ holds,} \end{cases} \quad \text{and} \quad \sigma^2 = \begin{cases} \sigma_0^2 \approx \sum_{i=0}^{\nu} p_i \left(\log(p_i/\theta_i)\right)^2 - \mu_0^2 & \text{if } H_0 \text{ holds;} \\ \sigma_1^2 \approx \sum_{i=0}^{\nu} \theta_i \left(\log(\theta_i/p_i)\right)^2 - \mu_1^2 & \text{if } H_1 \text{ holds.} \end{cases}$$

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 (30) is applied with T_κ as the test statistic. Then

$$\begin{aligned} \alpha &= \Pr[\text{Type-I error}] = \Pr\left[T_\kappa \le t \,|H_0 \text{ holds}\right] = \Phi\left(\frac{t - N\mu_0}{\sigma_0\sqrt{N}}\right) = 1 - P_S; \\ \beta &= \Pr\left[T_\kappa > t \,|H_1 \text{ holds}\right] = 1 - \Phi\left(\frac{t - N\mu_1}{\sigma_1\sqrt{N}}\right). \end{aligned}$$

Eliminating t between these two expressions and using (32) to substitute 2^{-a} for β gives the following expression for N.

$$N = \frac{\left\{\sigma_1 \Phi^{-1}\left(1 - 2^{-a}\right) + \sigma_0 \Phi^{-1}\left(P_S\right)\right\}^2}{\left(\mu_0 - \mu_1\right)^2}. \tag{56}$$

This expression is identical to the expression found by Blondeau et al in [11].

7.4 The χ^2 Test Statistic

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

$$T_\kappa = N \sum_{i=0}^{\nu} \frac{(X_{\kappa,i}/N - \theta_i)^2}{\theta_i}.$$

The hypothesis testing framework of (30) is applied.

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

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.

$$N = \frac{1}{C(\tilde{p})} \left[2\left(\Phi^{-1}(P_S)\right)^2 + \Psi^{-1}\left(1 - 2^{-a}\right) - \nu + \Phi^{-1}(P_S) \sqrt{4\left(\Phi^{-1}(P_S)\right)^2 + 2\Psi^{-1}\left(1 - 2^{-a}\right)} \right]; \qquad (57)$$

where, Ψ is the cumulative distribution function of the χ^2 distribution. This expression is identical to the expression found by Blondeau et al in [11].

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 [20, 11] and the corresponding normal approximations. In [20], the use was justified by a reference to the work by Drost et al [16]. On checking [16] 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 that those we obtain.

8.1 From Bernoulli to Normal to χ^2 to Normal

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

$$p_{\min} = \min_{\eta} p_\eta. \qquad (58)$$

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 $X_1, X_2, \ldots, X_\lambda$ be independent random variables each following $\tilde{p}$. We next describe a sequence of transformations on the X_j's 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

$$Q_{\eta,j} = \begin{cases} 1 & \text{if } X_j = \eta; \\ 0 & \text{otherwise.} \end{cases} \qquad (59)$$

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

From Bernoulli to Normal: For $1 \leq \eta \leq k$, define

$$Q_\eta \;=\; \sum_{j=1}^{\lambda} Q_{\eta,j}. \tag{60}$$

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 1 provides the condition on λ for which such approximation can be considered to be good. Denote the bound on the right side of Corollary 1 by Λ_η. This depends on η and it is necessary to consider the maximum of all the Λ_η's. Define

$$\Lambda\left(\tilde{p}\right) \;\triangleq\; \max_{\eta} \Lambda_\eta, \text{ where } \Lambda_\eta = \frac{C\left(p_\eta^2 + (1-p_\eta)^2\right)}{\left(p_\eta\left(1-p_\eta\right)\right)^{1/2}}. \tag{61}$$

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

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

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 $\mathcal{N}\left(\lambda p_\eta, \lambda p_\eta\left(1-p_\eta\right)\right)$. Note that we have $(k-1)$ independently distributed random variables Q_η, $\eta = 1, 2, \ldots, k-1$ and one dependent random variable Q_k given by $Q_\eta = \lambda - \sum_{\eta=1}^{k-1} Q_\eta$. Let,

$$Y \;=\; \sum_{\eta=1}^{k} \frac{\left(Q_\eta - \lambda\pi_\eta\right)^2}{\lambda\pi_\eta(1-\pi_\eta)} = \lambda \times \sum_{\eta=1}^{k} \frac{\left(Q_\eta/\lambda - \pi_\eta\right)^2}{\pi_\eta(1-\pi_\eta)}. \tag{62}$$

where $\tilde{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)$ 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 Chapter 18 of [21]) 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,

$$\begin{aligned}
Y &= \sum_{\eta=1}^{k} \frac{\left(Q_\eta - \lambda\pi_\eta\right)^2}{\lambda\pi_\eta(1-\pi_\eta)} = \sum_{\eta=1}^{k} \frac{\left(Q_\eta - \lambda\pi_\eta\right)^2}{\lambda p_\eta(1-p_\eta)} \times \left\{1 - \frac{p_\eta(1-p_\eta) - \pi_\eta(1-\pi_\eta)}{p_\eta(1-p_\eta)}\right\}^{-1} \\
&= \sum_{\eta=1}^{k} \frac{\left(Q_\eta - \lambda\pi_\eta\right)^2}{\lambda p_\eta(1-p_\eta)} \times \left\{1 + \frac{p_\eta(1-p_\eta) - \pi_\eta(1-\pi_\eta)}{p_\eta(1-p_\eta)} + \left(\frac{p_\eta(1-p_\eta) - \pi_\eta(1-\pi_\eta)}{p_\eta(1-p_\eta)}\right)^2 + \cdots\right\} \\
&\approx \sum_{\eta=1}^{k} \frac{\left(Q_\eta - \lambda\pi_\eta\right)^2}{\lambda p_\eta(1-p_\eta)}.
\end{aligned} \tag{63}$$

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 $(p_\eta(1-p_\eta) - \pi_\eta(1-\pi_\eta))/(p_\eta(1-$

$p_\eta)) \approx 0$. 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 Chapter 29.1 of [22]).

$$Y \sim \chi^2_{k-1}(\lambda D_{\tilde{p},\tilde{\pi}}) \text{ where } D_{\tilde{p},\tilde{\pi}} = \sum_{\eta=1}^{k} \frac{(p_\eta - \pi_\eta)^2}{p_\eta(1-p_\eta)}. \tag{64}$$

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

$$D_{\tilde{p},\tilde{\pi}} \approx C_{\tilde{p},\tilde{\pi}} \tag{65}$$

where

$$C_{\tilde{p},\tilde{\pi}} = \sum_{\eta=1}^{k} \frac{(p_\eta - \pi_\eta)^2}{\pi_\eta(1-\pi_\eta)}. \tag{66}$$

The conditions required for the Taylor series expansion and the approximation (65) to hold are $|p_\eta(1-p_\eta) - \pi_\eta(1-\pi_\eta)| < \pi_\eta(1-\pi_\eta)$ and $(p_\eta(1-p_\eta) - \pi_\eta(1-\pi_\eta))/(\pi_\eta(1-\pi_\eta)) \approx 0$.

All the required conditions for the approximations (63) and (65) 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 (64) under the approximation given by (65). Define $k-1$ independent and identically distributed random variables Y_i such that $Y_i \sim \mathcal{N}\left(\sqrt{\frac{\lambda C_{\tilde{p},\tilde{\pi}}}{k-1}}, 1\right)$ and $Y_i^2 \sim \chi^2_1\left(\frac{\lambda C_{\tilde{p},\tilde{\pi}}}{k-1}\right)$, $i = 1, 2, \ldots, k-1$ (see "Characterization" under Chapter 29.5 of [22]). It is known (see "Reproductivity" under Chapter 29.5 of [22]) that $\sum_{i=1}^{k-1} Y_i^2$ follows $\chi^2_{k-1}(\lambda C_{\tilde{p},\tilde{\pi}})$ 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.

Each $Y_i^2 \sim \chi^2_1\left(\frac{\lambda C_{\tilde{p},\tilde{\pi}}}{k-1}\right)$ and so $E[Y_i^2] = 1 + (\lambda C_{\tilde{p},\tilde{\pi}})/(k-1)$. Let, $Z_i = Y_i^2 - \left(1 + \frac{\lambda C_{\tilde{p},\tilde{\pi}}}{k-1}\right)$. Then, $E[Z_i] = 0$;

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

$$\begin{aligned}
\rho &= E\left[|Z_i|^3\right] \le E\left[\left(Z_i^2\right)^3\right] + \left(1 + \frac{\lambda C_{\tilde{p},\tilde{\pi}}}{k-1}\right)^3; \qquad \text{[Since, } Y_i^2 > 0, \left(1 + \tfrac{\lambda C_{\tilde{p},\tilde{\pi}}}{k-1}\right) > 0\text{]} \\
&= 2\left(1 + \frac{\lambda C_{\tilde{p},\tilde{\pi}}}{k-1}\right)^3 + 6\left(1 + \frac{\lambda C_{\tilde{p},\tilde{\pi}}}{k-1}\right)\left(1 + 2\frac{\lambda C_{\tilde{p},\tilde{\pi}}}{k-1}\right) + 8\left(1 + 3\frac{\lambda C_{\tilde{p},\tilde{\pi}}}{k-1}\right) \\
&= \frac{1}{(k-1)^3}\left\{2\left(k - 1 + \lambda C_{\tilde{p},\tilde{\pi}}\right)^3 + 6\left(k-1\right)\left(k - 1 + \lambda C_{\tilde{p},\tilde{\pi}}\right)\left(k - 1 + 2\lambda C_{\tilde{p},\tilde{\pi}}\right) + 8\left(k-1\right)^2\left(k - 1 + 3\lambda C_{\tilde{p},\tilde{\pi}}\right)\right\} \\
&= \rho' \text{ (say).}
\end{aligned}$$

Note that for $C_{\tilde{p},\tilde{\pi}} = 0$, $\sigma^2 = 2$ and $\rho' = 16$. More generally, $\rho \leq \rho' < \infty$. So, Theorem 1 applies and we have

$$\begin{aligned}\left|\Pr\left[\frac{\sum_{i=1}^{k-1} Z_i}{\sqrt{k-1}\sigma} < x\right] - \Phi(x)\right| &= \left|\Pr\left[\frac{\sum_{i=1}^{k-1} Y_i^2 - (k-1+\lambda C_{\tilde{p},\tilde{\pi}})}{\sqrt{k-1}\sigma} < x\right] - \Phi(x)\right| \\ &= \left|\Pr\left[\frac{Y - (k-1+\lambda C_{\tilde{p},\tilde{\pi}})}{\sqrt{k-1}\sigma} < x\right] - \Phi(x)\right| \\ &\leq \frac{C\rho}{\sigma^3\sqrt{k-1}} \leq \frac{C\rho'}{\sigma^3\sqrt{k-1}}. \end{aligned} \tag{67}$$

If the bound on the right side of (67) 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 (67) 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; $\tilde{\pi}$ is the uniform distribution over $\{0,1\}^\ell$. Under these conditions, the random variable Y given by (62) is not the same as the test statistic T_κ given in (39). On the other hand, if we make the additional assumption that $1 - \pi_\eta \approx 1$, then Y given by (62) becomes identical to T_κ given in (39). 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 (39) to be equal to Y given by (62): $1 - 2^{-\ell} \approx 1$.

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 (40). 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.

1. **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 a surprising 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.

2. **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.
3. **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.

8.4 The Multinomial Approach

The tuple $(Q_1, \ldots, Q_k)$ defined using (60) 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 Theorem 3 of [32]. 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 $O(k^{1/2})$ [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 throws up a surprising consequence.

Recall from (46) the LLR test statistics $T_{\kappa,z} = \sum_{\eta\in\{0,1\}^\ell} Q_{\kappa,z,\eta} \log\left(\tilde{p}_z(\eta)/2^{-\ell}\right)$. This can be written in terms of $X_{\kappa,z,j}$, $j = 1,\ldots,N$ as follows:

$$T_{\kappa,z} = \sum_{\eta\in\{0,1\}^\ell} Q_{\kappa,z,\eta} \times \log\left(\frac{\tilde{p}_z(\eta)}{2^{-\ell}}\right) \tag{68}$$

$$= \sum_{j=1}^{N} \log\left(\frac{\tilde{p}_z\left(X_{\kappa,z,j}\right)}{2^{-\ell}}\right). \tag{69}$$

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

Let $\mu = E[Y_j]$, $\sigma^2 = E[(Y_j-\mu)^2]$ and $\rho = E[|Y_j-\mu|^3]$. Since the Y_j's take values from a finite set, clearly $\rho < \infty$ and so Theorem 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 $C\rho/(\sigma^3 N)^{1/2}$. For this upper bound to be at most $2^{-\varepsilon}$, N has to satisfy

$$N \geq \frac{C^2\rho^2 2^{2\varepsilon}}{\sigma^3}. \tag{70}$$

This gives a lower bound on N which needs to be contrasted with the expression for N given by (48). Approximate expressions for μ and σ^2 are derived next, but, deriving an approximation for ρ seems to be difficult. This makes it difficult to compare the expression given by (70) to that given by (48).

Using results on the Taylor series expansion of the Kullback-Liebler divergence from [14] (see also [2]), it was mentioned in [20] that if all the ϵ_η's are small, then $T_{\kappa,z}$ asymptotically follows $\mathcal{N}\left(N\mu, N\sigma^2\right)$, where

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

Here $D(\tilde{p}_z||\tilde{p}_\$)$ is the Kullback-Leibler divergence. The approximation is done as follows.

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

For each η, the Taylor series expansion of the logarithmic term is done to obtain the approximations of μ_0, μ_1, σ_0^2 and σ_1^2 in terms of the capacity. These 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 $2^{-\ell} \leq \epsilon_\eta \leq 1-2^{-\ell}$ 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 LLR based approach is inapplicable and cannot be used to justify the expression for the data complexity.

Consider for example that for some η, $\epsilon_\eta = 2^{-\ell}$. 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-Shamir [28] has shown 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 LLR based approach will not be applicable to such a situation.

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] On the Distribution of Linear Biases: Three Instructive Examples, author=Abdelraheem, Mohamed Ahmed and Ågren, Martin and Beelen, Peter and Leander, Gregor, booktitle=Advances in Cryptology–CRYPTO 2012, pages=50–67, year=2012, publisher=Springer.

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

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

[4] V Bentkus. Dependence of the Berry-Esseen Estimate on the Dimension. *Lithuanian Mathematical Journal*, 26(2):110–114, 1986.

[5] Andrew C. Berry. The Accuracy of the Gaussian Approximation to the Sum of Independent Variates. *Transactions of the American Mathematical Society*, 49(1):pp. 122–136, 1941.

[6] Eli Biham and Adi Shamir. Differential Cryptanalysis of DES-like Cryptosystems. In *Advances in Cryptology–CRYPTO'90*, pages 2–21. Springer, 1990.

[7] Eli Biham and Adi Shamir. Differential Cryptanalysis of DES-like Cryptosystems. *Journal of CRYPTOLOGY*, 4(1):3–72, 1991.

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

[9] Céline Blondeau, Andrey Bogdanov, and Gregor Leander. Bounds in Shallows and in Miseries. In *Advances in Cryptology–CRYPTO 2013*, pages 204–221. Springer, 2013.

[10] Céline Blondeau and Benoît Gérard. Multiple Differential Cryptanalysis: Theory and Practice. In *Fast Software Encryption*, pages 35–54. Springer, 2011.

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

[12] Céline Blondeau, Benoît Gérard, and Jean-Pierre Tillich. Accurate Estimates of the Data Complexity and Success Probability for Various Cryptanalyses. *Designs, Codes and Cryptography*, 59(1-3):3–34, 2011.

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

[14] Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. John Wiley & Sons, second edition, 2012.

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

[16] FC Drost, WCM Kallenberg, DS Moore, and J Oosterhoff. Power Approximations to Multinomial Tests of Fit. *Journal of the American Statistical Association*, 84(405):130–141, 1989.

[17] Carl-Gustaf Esseen. On the Liapounoff Limit of Error in the Theory of Probability. *Arkiv för matematik, astronomi och fysik*, 1942(A28):1–19, 1942.

[18] Carl-Gustav Esseen. A Moment Inequality with an Application to the Central Limit Theorem. *Scandinavian Actuarial Journal*, 1956(2):160–170, 1956.

[19] Willliam Feller. *An Introduction to Probability Theory and Its Applications, Volume 2*. John Wiley & Sons, 2008.

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

[21] Norman Lloyd Johnson, Samuel Kotz, and Narayanaswamy Balakrishnan. *Continuous Univariate Distributions, Volume 1*. Wiley Series in Probability and Statistics. John Wiley & Sons, second edition, 1994.

[22] Norman Lloyd Johnson, Samuel Kotz, and Narayanaswamy Balakrishnan. *Continuous Univariate Distributions, Volume 2*. Wiley Series in Probability and Statistics. John Wiley & Sons, second edition, 1995.

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

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

[25] Burton S Kaliski Jr and Matthew JB Robshaw. Linear Cryptanalysis Using Multiple Approximations. In *Advances in Cryptology–Crypto94*, pages 26–39. Springer, 1994.

[26] Lars R Knudsen. Truncated and Higher Order Differentials. In *Fast Software Encryption*, pages 196–211. Springer, 1995.

[27] Gregor Leander. On linear hulls, statistical saturation attacks, present and a cryptanalysis of puffin. In *Advances in Cryptology–EUROCRYPT 2011*, pages 303–322. Springer, 2011.

[28] Itsik Mantin and Adi Shamir. A Practical Attack on Broadcast RC4. In *Fast Software Encryption*, pages 152–164. Springer, 2002.

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

[30] Mitsuru Matsui. The First Experimental Cryptanalysis of the Data Encryption Standard. In Y. G. Desmedt, editor, *Advances in Cryptology–Crypto94*, pages 1–11. Springer, 1994.

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

[32] VV Sazonov. On the Multi-Dimensional Central Limit Theorem. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 181–204, 1968.

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

[34] Ilya S Tyurin. An Improvement of Upper Estimates of the Constants in the Lyapunov Theorem. *Russian Mathematical Surveys*, 65(3):201–202, 2010.

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