Noname manuscript No.
(will be inserted by the editor)

Accurate Estimates of the Data Complexity and Success Probability for Various Cryptanalyses

Céline Blondeau · Benoît Gérard · Jean-Pierre Tillich

the date of receipt and acceptance should be inserted later

Abstract Many attacks on encryption schemes rely on statistical considerations using plaintext/ciphertext pairs to find some information on the key. We provide here simple formulae for estimating the data complexity and the success probability which can be applied to a lot of different scenarios (differential cryptanalysis, linear cryptanalysis, truncated differential cryptanalysis, etc.). Our work does not rely here on Gaussian approximation which is not valid in every setting but use instead a simple and general approximation of the binomial distribution and asymptotic expansions of the beta distribution.

Keywords statistical cryptanalysis · success probability · data complexity

1 Introduction

Statistical attacks against ciphers aim at determining some information on the key. Such attacks rely on the fact that some phenomenon occurs with different probabilities depending on the key. Here we focus on the case where the attacker has a certain amount of plaintext/ciphertext pairs from which he extracts, for each possible key, N binary samples whose sum follows a binomial distribution of parameters (N, p_0) in the case of the good key and (N, p) otherwise. Such attacks are referred as *non-adaptive iterated attacks* by Vaudenay [1]. The problem addressed by all these attacks is to determine whether the sample results from a binomial distribution of parameters p_0 or p. The variety of statistical attacks covers a huge number of possibilities for (p_0, p). For instance, in linear cryptanalysis [2–4], p_0 is close to $p = \frac{1}{2}$ while in differential cryptanalysis [5], p is small and p_0 is quite larger than p.

In order to compare these attacks, the success probability must be evaluated. It is crucial to determine how this quantity behaves in terms of data

INRIA project-team SECRET, France
E-mail: {celine.blondeau, benoit.gerard, jean-pierre.tillich}@inria.fr

complexity (or the other way round how the data complexity depends on the success probability). To achieve this, it is necessary to have accurate estimates for the tails of binomial distributions.

This kind of work has already been done for differential and linear cryptanalysis. A normal approximation of the binomial law provides formulae of the success probability [6] and the data complexity [3] in the case of linear cryptanalysis. For differential cryptanalysis a well-known formula of the data complexity is obtained using a Poisson approximation for the binomial law [8]. To the best of our knowledge, no explicit formulae of the data complexity and success probability are given for other types of statistical cryptanalyses such as truncated differential attack [9] for instance.

1.1 Related work

The difficulty in finding the data complexity comes from the fact that the binomial law is not easy to handle in the cryptanalytic range of parameters. Ideally, we would like to have an approximation that can be used on the whole space of parameters. Actually, binomial tails vary with the number of samples N as a product of a polynomial factor $Q(N)$ and an exponential factor $e^{-\Gamma N}$:

$$Q(N)e^{-\Gamma N}. \tag{1}$$

The asymptotic behavior of the exponent has been exhibited by Baignères, Junod and Vaudenay [10–12] by applying some classical results from statistics. However, for many statistical cryptanalyses, the polynomial factor is non negligible. As far as we know, all previous works give estimates of this value using a Gaussian approximation that recovers the right polynomial factor but with an exponent which is only valid in a small range of parameters. For instance, the deep analysis of the complexity of linear attacks due to Junod [13, 10, 14] is based on a Gaussian approximation and cannot be adapted directly to other scenarios, like the different variants of differential cryptanalysis.

1.2 A practical instance: comparing truncated differential and differential attacks

The initial problem we wanted to solve was to compare the data complexity of a truncated differential attack and a differential attack. In a truncated differential cryptanalysis the probabilities p_0 and p are slightly larger than in a differential cryptanalysis but the ratio p_0/p is closer to 1.

Definition 1 Let F be a function with input space X and output space Y. A truncated differential for F is a pair of subsets (A, B), $A \subset X$, $B \subset Y$. The probability of this truncated differential is:

$$P_{x \in X}\left[F(x) + F(x+a) \in B | a \in A\right].$$

Hereafter we present both attacks on generalized Feistel networks [15] defined in Appendix 7.1. As a toy example, we study a generalized Feistel network with four S-boxes and ten rounds. The S-boxes are all the same and are defined over the field $GF(2^8)$ by the power permutation $x \mapsto x^7$.

Let T be a partition of $GF(2^8)$ into cosets of the subfield $GF(2^4)$. If η is a generator of $GF(2^8)$ with minimal polynomial $x^8 + x^4 + x^3 + x^2 + 1$, we define two cosets of $GF(2^4)$ by $T_1 = \eta^7 + GF(2^4)$ and $T_2 = GF(2^4)$. Let

$$A = (T_1, 0, 0, 0, 0, 0, 0, 0) \quad \text{and} \quad B = (T_1, T_2, ?, ?, ?, ?, T_1, T_2).$$

Note that A is the set of vectors of the form $(a, 0, 0, 0, 0, 0, 0, 0)$ where $a \in T_1$. Also '?' in B means any elements of $GF(2^8)$ (see Appendix 7.1).
For ten rounds of this generalized Feistel network with good subkeys, the probability of the truncated differential characterized by (A, B) is

$$p_0 = 1.18 \times 2^{-16}.$$

For the wrong subkeys, the output difference is supposed to be independent from the input difference. Thus, the probability for the output to be in B is :

$$p = \left(2^4/2^8\right)^4 = 2^{-16}.$$

The best differential cryptanalysis is derived from the same characteristic but with T_1 and T_2 reduced to one element ($T_1 = \{\alpha^{85}\}$ and $T_2 = \{0\}$). In this case, we have

$$p_0 = 1.53 \times 2^{-27} \quad \text{and} \quad p = (1/2^8)^4 = 2^{-32}.$$

The problem is then to determine whether the data complexity of the truncated differential cryptanalysis is lower than the data complexity of the differential cryptanalysis or not.

1.3 Our contribution

The main difficulty in expressing the data complexity comes from the fact that the binomial tails are not easy to handle. In this paper we use a simple approximation that is valid over a wide range of parameters. The approximation catches the right behavior of the polynomial term and the right exponential term as well as in (1).
We will compute the amount of data which is needed or the success probability in terms of the data complexity in two different scenarios:

(i) when the probability β that a wrong key is accepted is fixed,
(ii) when the size of the list of the kept candidates is fixed.

To simplify the expressions in scenario (i), we fix the success probability to 50% and give, in this case, an accurate estimate of the data complexity in terms of β. We also provide an asymptotic expression of this quantity for several types of cryptanalyses. Then we study scenario (ii) and provide a generalization of the formula of Selçuk [6] which gave an accurate expression for the probability of success only in the case of linear cryptanalysis. This formula relies heavily on Gaussian approximations which are not valid anymore in the case of differential cryptanalysis. On the contrary, the generalization presented in this paper is obtained using the aforementioned approximation of the binomial tail and asymptotic expansion of the tail of the beta distribution. Our formula gives an accurate expression which is valid for various cryptanalyses (and thus including differential cryptanalysis, truncated differential cryptanalysis and linear cryptanalysis).

2 Statistical cryptanalysis

The core of a statistical cryptanalysis is to use some statistical phenomenon to extract some information on the key used to encipher the intercepted plaintext/ciphertext pairs. We denote by N the number of available samples of plaintext/ciphertext. A sample can be composed of one pair (linear cryptanalysis), two pairs with chosen plaintexts (differential cryptanalysis), etc. Generally, the observed phenomenon only gives information on a subkey of the master key. Such attacks basically consists in three steps:

- *Distillation phase:* some statistic Σ is extracted from the available data.
- *Analysis phase:* from Σ, the likelihood of each possible subkey is computed and a list $\mathcal{L}$ of the likeliest keys is suggested.
- *Search phase:* for each subkey in $\mathcal{L}$, all the possible corresponding master keys are exhaustively tried until the good one is found.

We denote by K the random variable corresponding to the correct subkey. The likelihood of a subkey k is then $P\left[K = k|\Sigma\right]$. In most of the case, the statistic Σ is a set of counters Σ_k that correspond to the number of times some phenomenon (called characteristic) is observed for a subkey k. For a fixed subkey k, let X_k^i be a random variable that takes value 1 if the characteristic appears in the sample number i and takes value 0 otherwise. Thus, $X_k^1, .., X_k^N$ are N binary random variables which are independent and identically distributed. The counter Σ_k then corresponds to the sum of the X_k^i's:

$$\Sigma_k \overset{\text{def}}{=} \sum_{i=1}^{N} X_k^i \; .$$

We denote by p_0 the probability that the characteristic is observed when k is the correct subkey k_0:

$$p_0 \overset{\text{def}}{=} P(X_{k_0}^1 = 1) = \cdots = P(X_{k_0}^N = 1).$$

We assume that the Wrong-Key Randomization Hypothesis holds [7]: the phenomenon is observed with the same probability p independently of the value of the wrong key k:

$$p \overset{\text{def}}{=} P(X^1_{k \neq k_0} = 1) = \cdots = P(X^N_{k \neq k_0} = 1).$$

The counter Σ_k thus follows a binomial law with parameters (N, p_0) if k is the correct subkey and (N, p) otherwise $(p < p_0)$. In our setting, likelihoods are directly linked to counters Σ_k. Let k and k' be two subkeys,

$$P\left[K = k|\Sigma\right] \leq P\left[K = k'|\Sigma\right] \Longleftrightarrow \Sigma_k \leq \Sigma_{k'}.$$

This is actually the case for standard statistical cryptanalyses. From this setting, two paradigms can be studied. The first one is to fix some threshold and to accept in $\mathcal{L}$ all the subkeys with a likelihood more than this threshold. The second one is to fix the size of $\mathcal{L}$ to some integer ℓ and then keep the ℓ likeliest subkeys. These two paradigms are studied in the following two subsections.

2.1 Hypothesis Testing

Here we deal with the hypothesis testing paradigm. The problematic consists in fixing a threshold T and comparing the counter Σ_k with T:

$$\text{If } \Sigma_k \geq T \text{ then } k \in \mathcal{L} \text{ else } k \notin \mathcal{L}.$$

From the N samples, the attacker either decides that $k = k_0$ holds or that $k \neq k_0$ is true. Two kinds of errors are possible:

- **Non-detection:** It occurs if one decides that $k \notin \mathcal{L}$ when $k = k_0$ holds. We denote by α the non-detection error probability.
- **False alarm:** It occurs if one decides that $k \in \mathcal{L}$ when $k \neq k_0$ holds. We denote by β the false alarm error probability.

Using well known results about hypothesis testing it follows that, for some integer $0 \leq T \leq N$, $\{\Sigma_k\ ,\ \Sigma_k \geq T\}$ is an optimal acceptance region. The meaning of optimal is stated in the following lemma.

Lemma 1 *[16]**Neyman-Pearson lemma** :*
If distinguishing between two hypotheses $k = k_0$ and $k \neq k_0$ with the help of N variables $(X^i_k)_i$ and using a test of the form

$$\frac{P(X^1_k, \ldots, X^N_k | k = k_0)}{P(X^1_k, \ldots, X^N_k | k \neq k_0)} \geq t$$

gives error probabilities α and β, then no other test can improve both non-detection and false alarm error probabilities.

A standard calculus (detailed in [16] for the Gaussian case) shows that comparing the ratio of Lemma 1 with a real number t is equivalent to compare Σ_k with an integer $0 \leq T \leq N$.

2.2 Key ranking

Here we deal with the key ranking paradigm. The problematic is not to decide if a subkey is the good one or not but to distinguish the correct subkey from many incorrect ones. We denote by n the total number of possible subkeys: the correct one k_0 plus $n-1$ incorrect subkeys $k_1, \ldots, k_{n-1}$. Then, the idea is to keep a list $\mathcal{L}$ of the ℓ subkeys that are the more likely to be the correct one.

$$\forall k \notin \mathcal{L} \ , \ \forall k' \in \mathcal{L} \ , \varSigma_k \leq \varSigma_{k'}$$

The cryptanalysis is a success if the correct key belongs to this list.

Definition 2 We define the *success probability* of a statistical cryptanalysis as the probability that the correct subkey k_0 belongs to the list of the ℓ likeliest subkeys.

$$P_S \overset{\text{def}}{=} P\left[k_0 \in \mathcal{L}\right].$$

The following section give estimates for the error probability in order to find a simple expression of the data complexity. We will go back to the key ranking problem in Section 5 where we give a simple formula to estimate P_S.

3 Approximating error probabilities

3.1 The binomial distribution

Since it is difficult to handle the binomial law, we need to approximate it. A particular quantity will play a fundamental role here, the Kullback-Leibler divergence.

Definition 3 Kullback-Leibler divergence [16]
Let $\mathcal{P}$ and $\mathcal{Q}$ be two Bernoulli probability distributions of respective parameters p and q. The Kullback-Leibler divergence between $\mathcal{P}$ and $\mathcal{Q}$ is defined by:

$$D\left(p||q\right) \overset{\text{def}}{=} p \ln\left(\frac{p}{q}\right) + (1-p)\ln\left(\frac{1-p}{1-q}\right).$$

We use the convention (based on continuity arguments) that $0 \ln \frac{0}{p} = 0$ and $p \ln \frac{p}{0} = \infty$.

Lemma 2 *Let τ be a relative threshold $0 \leq \tau \leq 1$. Let $\varSigma_k$ be a random variable that follows a binomial law of parameters (N, p). We have:*

$$P(\varSigma_k = \lfloor \tau N \rfloor) = \sqrt{\frac{1}{2\pi N(1-\tau)\tau}}\, e^{-ND(\tau||p)} \left[1 + O\left(\frac{1}{\tau N}\right)\right]. \tag{2}$$

Proof We recall the probability function of the binomial law:

$$P(\Sigma_k = \lfloor \tau N \rfloor) = \binom{N}{\lfloor \tau N \rfloor} p^{\lfloor \tau N \rfloor}(1-p)^{N-\lfloor \tau N \rfloor}.$$

Using the Stirling approximation we have

$$\binom{N}{\lfloor \tau N \rfloor} = \sqrt{\frac{1}{2\pi N\tau(1-\tau)}}\; e^{-N[\tau\ln(\tau)-(1-\tau)\ln(1-\tau)]}\left[1+O\left(\frac{1}{\tau N}\right)\right]$$

and writing

$$p^{\tau N}(1-p)^{N-\tau N} = e^{\tau N\ln(p)+(N-\tau N)\ln(1-p)},$$

we obtain

$$P(\Sigma_k = \lfloor \tau N \rfloor) = \sqrt{\frac{1}{2\pi\tau(1-\tau)}} \cdot e^{-N\left[\tau\ln\left(\frac{\tau}{p}\right)+(1-\tau)\ln\left(\frac{1-\tau}{1-p}\right)\right]}\left[1+O\left(\frac{1}{\tau N}\right)\right].$$

□

Lemma 3 *Let Σ_k be a random variable that follows a binomial law of parameters (N, p). Let A and B be two integers such that $0 \le A \le B \le N$. Let $\gamma_+ \overset{\text{def}}{=} \frac{1-p}{p}\max\left(\frac{B}{N-B+1}, \frac{A+1}{N-A}\right)$ and $\gamma_- \overset{\text{def}}{=} \frac{1-p}{p}\min\left(\frac{B}{N-B+1}, \frac{A+1}{N-A}\right)$. Then, we have*

$$P(\Sigma_k = B)\frac{1-\gamma_-^{B-A+1}}{1-\gamma_-} \le \sum_{i=A}^{B} P(\Sigma_k = i) \le P(\Sigma_k = B)\frac{1-\gamma_+^{B-A+1}}{1-\gamma_+},$$

$$P(\Sigma_k = A)\frac{1-1/\gamma_+^{B-A+1}}{1-1/\gamma_+} \le \sum_{i=A}^{B} P(\Sigma_k = i) \le P(\Sigma_k = A)\frac{1-1/\gamma_-^{B-A+1}}{1-1/\gamma_-},$$

Proof We can see that

$$P(\Sigma_k = i-1) = \frac{1-p}{p}\frac{i}{N-i+1}P(\Sigma_k = i), \text{ for } 0 < i \le N.$$

This leads to:

$$\sum_{i=A}^{B} P(\Sigma_k = i) = P(\Sigma_k = B)\left[1+\frac{(1-p)B}{p(N-B+1)}+\cdots+\frac{(1-p)^{B-A}B\cdots(A+1)}{p^{B-A}(N-B+1)\cdots(N-A)}\right].$$

We deduce that

$$P(\Sigma_k = B)\sum_{i=0}^{B-A}\gamma_-^i \le \sum_{i=A}^{B} P(\Sigma_k = i) \le P(\Sigma_k = B)\sum_{i=0}^{B-A}\gamma_+^i.$$

This implies the lemma.

□

Notation 1 Writing $f \underset{N\to\infty}{\sim} g$ means that $\lim\limits_{N\to\infty}\frac{f(N)}{g(N)} = 1$.

The next theorem is known in another context (see. [17]). We can derive for instance the first expression (3) from the previous lemmas by writing for $A \overset{\text{def}}{=} \lceil \tau N \rceil$ that $P(\Sigma_k \geq \tau N) = \sum_{i=A}^{N} P(\Sigma_k = i) = \sum_{i=A}^{B} P(\Sigma_k = i) + \sum_{i=B+1}^{N} P(\Sigma_k = i)$, and by applying Lemma 3 to the first sum and by choosing B such that
(i) the second sum is negligible in front of the first one,
(ii) and such that $\gamma_+ \approx \gamma_-$.

Theorem 1 *Let p_0 and p be two real numbers such that $0 < p < p_0 < 1$ and let τ such that $p < \tau < p_0$. Let Σ_k and Σ_0 follow a binomial law of respective parameters (N, p) and (N, p_0). Then,*

$$P(\Sigma_k \geq \tau N) \underset{N \to \infty}{\sim} \frac{(1-p)\sqrt{\tau}}{(\tau - p)\sqrt{2\pi N(1-\tau)}} e^{-ND(\tau || p)}, \tag{3}$$

and

$$P(\Sigma_0 \leq \tau N) \underset{N \to \infty}{\sim} \frac{p_0\sqrt{1-\tau}}{(p_0 - \tau)\sqrt{2\pi N\tau}} e^{-ND(\tau || p_0)}. \tag{4}$$

3.2 Comparison with other approximations

A formula valid in many cases. The approximation given in Theorem 1 is quite accurate over a very wide range of parameters (whether p is small or not whether τ is close to p or not). This is in sharp contrast with the approximations which have been used up to now. In the case of differential cryptanalysis where p is small and τ is significantly different from p, a Poisson approximation is used. It gives a sharp estimate but it is not valid anymore in the case of linear cryptanalysis where p is close to $1/2$ and τ is close to p. In this case, a Gaussian approximation is used instead, see [13,10,14,11,12,6]. However this Gaussian approximation gives poor estimates for differential cryptanalysis.

On the exponential behavior of the binomial tails. Binomial tails are well known to decrease exponentially in N. The correct exponent has been given in several places. For instance, in [11,12], the aim of the authors is to derive an asymptotic formula for the best distinguisher, that is the distinguisher that maximizes $|1 - \alpha - \beta|$. The following result is derived:

$$\max(\alpha, \beta) \doteq 2^{-NC(p_0, p)} \tag{5}$$

where $f(N) \doteq g(N)$ means $f(N) = g(N)e^{o(N)}$ and C is the Chernoff information.

In the general case where $p_0 \notin \{0, 1\}$, such a distinguisher has an acceptance region of the form mentioned by Lemma 1 with t equals to 1. In this setting, the value of the relative threshold τ fulfills the equality $D(\tau || p_0) = D(\tau || p)$. Actually, this value of the Kullback-Leibler divergence is equal to the

Chernoff information $C(p_0, p)$ times $\ln(2)$ (see [16, Section 12.9]). Thus, the exponent in (3) and (4) is the same as the one given by (5):

$$\alpha \doteq e^{-ND(\tau||p_0)} \doteq 2^{-NC(p_0,p)} \quad \text{and,} \quad \beta \doteq e^{-ND(\tau||p)} \doteq 2^{-NC(p_0,p)}.$$

In the case $p_0 = 0$ or $p_0 = 1$, in impossible or higher order differential cryptanalysis for instance, the relative threshold τ is equal to p_0 and the non-detection error probability α vanishes. Thus, $\max(\alpha, \beta) = \beta \doteq e^{-ND(p_0||p)} \doteq 2^{-NC(p_0,p)}$. The last equality is directly derived from the definition of the Kullback-Leibler divergence. The correct exponential behavior in $ND(\tau||p)$ is captured by our Theorem 1 but we also have an additional polynomial term $\frac{(1-p)\sqrt{\tau}}{(\tau-p)\sqrt{2\pi N(1-\tau)}}$ which is non negligible. Taking only the exponential term in (5) is too coarse in many cases as Figure 1 shows.
For instance in Figure 1, the data complexity given by the formula of [12] is four times larger than the real value N. Moreover, the lack of the polynomial term gives worse results when considering error probabilities. The error on the probabilities can be more than 50%.

On the polynomial behavior of the binomial tails. In [11], a polynomial factor is taken into account. However it is only suitable when the Gaussian approximation of binomial tails can be used. In this case, the data complexity is:

$$N \approx \frac{2 \cdot \Phi^{-1}(\frac{\alpha+\beta}{2})^2}{D(p_0||p)}, \tag{6}$$

where Φ^{-1} is the inverse cumulative function of a Gaussian random variable. For instance, this formula gives a poor estimate in the case of differential cryptanalysis. In general this formula is too optimistic as it can be seen in Figure 1.

Explanation of Figure 1. Hereafter we compare the real value of the data complexity N (the required number of samples) to the estimates obtained using (5) and (6). The value of $\log_2(N)$ is obtained thanks to Algorithm 1 presented in Subsection 4.1. An additional column contains the estimate found using (3) and (4). Notice that these estimate tends towards N as β goes to zero.

4 On the required Data Complexity

4.1 General method for finding the Data Complexity

We are interested in this section in finding an accurate number of samples to reach given error probabilities.

Let Σ_k (resp. Σ_0) be a random variable which follows a binomial law of parameters N and p (resp. p_0) in the hypothesis testing paradigm. The acceptance region is defined by a threshold T, thus both error probabilities

			$\log_2(N)$	(3) & (4)	[11]	[12]
Linear	$p_0 = 0.5 + 1.49 \cdot 2^{-24}$ $\alpha = 0.1$	$p = 0.5$ $\beta = 0.1$	47.57	47.88	47.57	49.58
Linear	$p_0 = 0.5 + 1.49 \cdot 2^{-24}$ $\alpha = 0.001$	$p = 0.5$ $\beta = 0.001$	50.10	50.13	50.10	51.17
Differential	$p_0 = 1.87 \cdot 2^{-56}$ $\alpha = 0.1$	$p = 2^{-64}$ $\beta = 0.1$	56.30	56.77	54.44	57.71
Differential	$p_0 = 1.87 \cdot 2^{-56}$ $\alpha = 0.001$	$p = 2^{-64}$ $\beta = 0.001$	58.30	58.50	56.98	59.29
Truncated differential	$p_0 = 1.18 \cdot 2^{-16}$ $\alpha = 0.001$	$p = 2^{-16}$ $\beta = 0.001$	26.32	26.35	26.28	27.39

Fig. 1 Comparison of estimates of $\log_2(N)$ from [11,12] and our work for various parameters.

can be rewritten as $P(\Sigma_0 < T)$ and $P(\Sigma_k \geq T)$. Let α and β be two given real numbers ($0 < \alpha, \beta < 1$). The problem is to find a number of samples N and a threshold T such that the error probabilities are less than α and β respectively. This is equivalent to find a solution (N, T) of the following system

$$\begin{cases} P(\Sigma_0 < T) \leq \alpha, \\ P(\Sigma_k \geq T) \leq \beta. \end{cases} \tag{7}$$

In practice, using real numbers avoids troubles coming from the fact that the set of integers is discrete. Thus, we use estimates on error probabilities that are functions with real entries N and $\tau = T/N$ (relative threshold). Formulae from Theorem 1 can be used for those estimates.

We denote respectively by $G_{\mathrm{nd}}(N, \tau)$ and $G_{\mathrm{fa}}(N, \tau)$ the estimates for non-detection and false alarm error probabilities. These estimates are chosen is such a way that they are decreasing functions in N for a given τ. In consequence, the problematic boils down to find N and τ such that

$$G_{\mathrm{nd}}(N, \tau) \leq \alpha \quad \text{and} \quad G_{\mathrm{fa}}(N, \tau) \leq \beta. \tag{8}$$

For a given τ, we compute the values $N_{\mathrm{nd}}(\tau)$ and $N_{\mathrm{fa}}(\tau)$ such that:

$$G_{\mathrm{nd}}(N_{\mathrm{nd}}(\tau), \tau) = \alpha \quad \text{and} \quad G_{\mathrm{fa}}(N_{\mathrm{fa}}(\tau), \tau) = \beta.$$

One of these two values may be greater than the other one. In this case, the threshold should be changed to balance N_{nd} and N_{fa}: for a fixed N, decreasing

τ means accepting more candidates and so non-detection error probability decreases while false alarm error probability increases.

Algorithm 1 then represents a method for computing the values of N and τ which correspond to balanced N_{fa} and N_{nd}. It is based on the following lemma.

Lemma 4 *Let $G_{\mathrm{nd}}(N,\tau)$ and $G_{\mathrm{fa}}(N,\tau)$ be two functions of N and τ, defined on $[0,+\infty) \times [p,p_0]$, with the following properties:*

- *for a fixed τ, both are decreasing functions of N;*
- *for a fixed N, $G_{\mathrm{nd}}(N,\tau)$ (resp. $G_{\mathrm{fa}}(N,\tau)$) is increasing (resp. decreasing) in τ;*
- $\lim_{N\to 0} G_{\mathrm{nd}}(N,\tau) \geq 1$, $\lim_{N\to 0} G_{\mathrm{fa}}(N,\tau) \geq 1$;
- $\lim_{N\to\infty} G_{\mathrm{nd}}(N,\tau) = \lim_{N\to\infty} G_{\mathrm{fa}}(N,\tau) = 0$.

Let us recall that for fixed α, β in $[0,1]$ and τ in $[p,p_0]$, $G_{\mathrm{nd}}(N_{\mathrm{nd}}(\tau),\tau) = \alpha$ and $G_{\mathrm{fa}}(N_{\mathrm{fa}}(\tau),\tau) = \beta$.
We introduce $N(\tau) = \max(N_{\mathrm{nd}}(\tau), N_{\mathrm{fa}}(\tau))$ which represents the minimal N such that (N,τ) fulfils (8).
Then, for $p \leq m \leq p_0$,
if $N_{\mathrm{nd}}(m) > N_{\mathrm{fa}}(m)$, then, for all $\tau > m$, $N(\tau) > N(m)$;
if $N_{\mathrm{nd}}(m) < N_{\mathrm{fa}}(m)$, then, for all $\tau < m$, $N(\tau) > N(m)$.

Proof Both proofs are similar, so we only prove the first statement. Since $N_{\mathrm{nd}}(m) > N_{\mathrm{fa}}(m)$, we have $G_{\mathrm{nd}}(N(m),m) = \alpha$ and $G_{\mathrm{fa}}(N(m),m) < \beta$. Using the increasing/decreasing properties of $G_{\mathrm{nd}}/G_{\mathrm{fa}}$, we can say that for $\tau > m$, $G_{\mathrm{nd}}(N(m),\tau) > \alpha$ and $G_{\mathrm{fa}}(N(m),\tau) < \beta$. Then, since those functions are decreasing with N, we deduce that $N(\tau) > N(m)$. □

Algorithm 1 Computation of the exact number of samples required for a statistical attack (and the corresponding relative threshold).

```
Input:   Given error probabilities $(\alpha,\beta)$ and probabilities $(p_0,p)$.
Output: $N$ and $\tau$: the minimum number of samples and the corresponding relative threshold to reach error probabilities less than $(\alpha,\beta)$.

Set $\tau_{min}$ to $p$ and $\tau_{max}$ to $p_0$.
repeat
  Set $\tau$ to $\frac{\tau_{min}+\tau_{max}}{2}$.
  Compute $N_{\mathrm{nd}}$ such that $\forall N > N_{\mathrm{nd}}, G_{\mathrm{nd}}(N,\tau) \leq \alpha$.
  Compute $N_{\mathrm{fa}}$ such that $\forall N > N_{\mathrm{fa}}, G_{\mathrm{fa}}(N,\tau) \leq \beta$.
  if $N_{\mathrm{nd}} > N_{\mathrm{fa}}$ then
    $\tau_{max} = \tau$.
  else
    $\tau_{min} = \tau$.
  end if
until $N_{\mathrm{nd}} = N_{\mathrm{fa}}$.
Return $N = N_{\mathrm{nd}} = N_{\mathrm{fa}}$ and $\tau$.
```

The computation of N_{nd} and N_{fa} can be made thanks to a dichotomic search but a more efficient way of doing that is explained in Appendix 7.2.

4.2 Asymptotic behaviour of the Data Complexity

The aim of this section is to provide a simple criterion to compare two different statistical attacks. An attack is defined by a pair (p_0, p) of probabilities where p (resp. p_0) is the probability that the phenomenon occurs for a wrong key output $k \neq k_0$ (resp. for the good key output $k = k_0$).

In order to simplify the following computation, we take a threshold $\tau = p_0$ that gives a non-detection error probability α of order $\frac{1}{2}$. In statistical attacks, the time complexity is related to the false alarm probability β, thus, it is important to control this probability. That is why taking $\tau = p_0$ is a natural way of simplifying the problem.

Then, we can use Theorem 1 to derive a sharp approximation of N introduced in the following theorem.

Theorem 2 *Let p_0 (resp. p) be the probability of the phenomenon to occur in the good key parametrization (resp. the wrong key parametrization). For a relative threshold $\tau = p_0$, a good approximation of the required number of samples N to distinguish between the correctly keyed permutation and a incorrectly keyed permutation with false alarm error probability less or equal to β is*

$$N' \overset{\text{def}}{=} -\frac{1}{D\left(p_0||p\right)}\left[\ln\left(\frac{\nu\beta}{\sqrt{D\left(p_0||p\right)}}\right) + 0.5\ln\left(-\ln(\nu\beta)\right)\right], \tag{9}$$

since

$$N' \leq N_\infty \leq N'\left[1 + \frac{(\theta - 1)\ln(\theta)}{\ln(N')}\right],$$

for

$$\nu \overset{\text{def}}{=} \frac{(p_0 - p)\sqrt{2\pi(1-p_0)}}{(1-p)\sqrt{p_0}} \quad \text{and} \quad \theta \overset{\text{def}}{=} \left[1 + \frac{1}{2\ln(\nu\beta)}\ln\left(-\frac{\ln(\nu\beta)}{D\left(p_0||p\right)}\right)\right]^{-1}. \tag{10}$$

Where N_∞ is the value obtained with Algorithm 1 using (3) and (4) as estimates of error probabilities.

Proof See Appendix 7.4. □

This approximation with N' is tight: we estimated the data complexity of some known attacks (see Figure 2) and observed θ's in the range $(1, 6.5]$. Moreover, for $\beta = 2^{-32}$, observed values of θ's were less than 2.

Equation (9) gives a simple way of roughly comparing the data complexity of two statistical attacks. Indeed, N' is essentially a decreasing function of

$D(p_0||p)$. Therefore, comparing the data complexity of two statistical cryptanalyses boils down to comparing the corresponding Kullback-Leibler divergences.

Moreover, it can be proved that $\ln(2\sqrt{\pi D(p_0||p)})$ is a good estimate of $\ln(\nu)$. Thus, a good approximation of N' is

$$N'' \overset{\text{def}}{=} -\frac{\ln(2\sqrt{\pi}\beta)}{D(p_0||p)}. \tag{11}$$

Experimental results given in Section 4.3 show that this estimation is quite sharp and becomes better as β goes to 0.

To have a more accurate comparison between two attacks (for instance in the case $\alpha \neq 0.5$), Algorithm 1 may be used. Notice that the results we give are estimates of the number of samples and not of the number of plaintexts. In the case of linear cryptanalysis it remains the same but in the case of differential, a sample is derived from a pair of plaintexts with a given differential characteristic. Thus, the number of required plaintexts is twice the number of samples. The estimate of the number of plaintexts is a more specific issue we will not deal with.

4.3 Experimental results

Here we present some results found with Algorithm 1 to show the accuracy of the estimate given by Theorem 2.

Let us denote by N the exact number of required samples, we want to compare it to both estimates. Let us write again both approximations of N given in Subsection 4.2, namely:

$$N' = -\frac{1}{D(p_0||p)}\left[\ln\left(\frac{\nu\beta}{\sqrt{D(p_0||p)}}\right) + 0.5\ln\left(-\ln(\nu\beta)\right)\right]$$

$$N'' = \frac{-\ln(2\sqrt{\pi}\beta)}{D(p_0||p)}.$$

In Figure 2, N is given with two decimal digits precision. This table compares the values of N' and N'' to the real value N for some parameters. We can see in Figure 2 that N' and N'' tend to N as β goes to 0.

4.4 Application on statistical attacks

Now that we have expressed N in terms of Kullback-Leibler divergence, we see that the behavior of N is dominated by $D(p_0||p)^{-1}$. Hereafter, we estimate $D(p_0||p)^{-1}$ for many statistical cryptanalyses. We recover the format of

		p	p_0	$\log_2(N)$	$\log_2(N')$	$\log_2(N'')$
$\beta = 2^{-8}$	**L**	0.5	$0.5 + 1.19 \cdot 2^{-21}$	42.32	42.00 (−0.32)	42.60
	DL	0.5	$0.5 + 1.73 \cdot 2^{-6}$	11.26	11.15 (−0.11)	11.52
	D	2^{-64}	$1.87 \cdot 2^{-56}$	54.57	54.68 (+0.11)	54.82
	Dgfn	2^{-32}	$1.53 \cdot 2^{-27}$	27.14	26.80 (−0.34)	26.94
	TDgfn	2^{-16}	$1.18 \cdot 2^{-16}$	23.85	23.66 (−0.19)	24.13
$\beta = 2^{-16}$	**L**	0.5	$0.5 + 1.19 \cdot 2^{-21}$	43.62	43.54 (−0.08)	43.79
	DL	0.5	$0.5 + 1.73 \cdot 2^{-6}$	12.54	12.52 (−0.02)	12.71
	D	2^{-64}	$1.87 \cdot 2^{-56}$	55.85	55.94 (+0.09)	56.02
	Dgfn	2^{-32}	$1.53 \cdot 2^{-27}$	28.27	28.05 (−0.22)	28.14
	TDgfn	2^{-16}	$1.18 \cdot 2^{-16}$	25.15	25.11 (−0.04)	25.33
$\beta = 2^{-32}$	**L**	0.5	$0.5 + 1.19 \cdot 2^{-21}$	44.78	44.76 (−0.02)	44.88
	DL	0.5	$0.5 + 1.73 \cdot 2^{-6}$	13.70	13.69 (−0.01)	13.80
	D	2^{-64}	$1.87 \cdot 2^{-56}$	56.98	57.06 (+0.08)	57.11
	Dgfn	2^{-32}	$1.53 \cdot 2^{-27}$	29.13	29.17 (+0.04)	29.23
	TDgfn	2^{-16}	$1.18 \cdot 2^{-16}$	26.31	26.30 (−0.01)	26.42

Fig. 2 Estimates and real value of the data complexity for som parameters β, p and p_0.

- **L** : DES linear cryptanalysis recovering 26 key bits [4].
- **DL** : DES differential-linear cryptanalysis [18].
- **D** : DES differential cryptanalysis [19].
- **Dgfn/TDgfn** : Generalized Feistel networks (truncated) differential cryptanalysis presented in this paper.

known results and give new results for truncated differential and higher order differential cryptanalysis. Let us recall the Kullback-Leibler divergence

$$D\left(p_0 || p\right) = p_0 \ln\left(\frac{p_0}{p}\right) + (1-p_0)\ln\left(\frac{1-p_0}{1-p}\right).$$

In Appendix 7.3, Lemma 7 gives an expansion of Kullback-Leibler divergence

$$D\left(p_0 || p\right) = p_0 \left[\log\left(\frac{p_0}{p}\right) - \frac{p_0 - p}{p_0} + \frac{(p_0-p)^2}{2p_0(1-p_0)}\right] + O(p_0-p)^3.$$

From this, we derive the asymptotic behavior of the number of sample for set of parameters depending of the type of cryptanalysis.

Attacks	Asymptotic behavior of the number of samples	Asymptotic behavior of the number of plaintexts	Known or chosen plaintexts (**CP/KP**)
Linear	$\frac{1}{2(p_0 - p)^2}$	$\frac{1}{2(p_0 - p)^2}$	**KP**
Differential	$\frac{1}{p_0 \ln(p_0/p) - p_0}$	$\frac{2}{p_0 \ln(p_0/p) - p_0}$	**CP**
Differential-linear	$\frac{1}{2(p_0 - p)^2}$	$\frac{1}{(p_0 - p)^2}$	**CP**
Truncated differential	$\frac{p}{(p_0 - p)^2}$	$\frac{p \cdot \gamma}{(p_0 - p)^2}$, $1 < \gamma < 2$	**CP**
Impossible differential	$\frac{1}{p}$	$\frac{2}{p}$	**CP**
i-th order differential	$-\frac{1}{\ln p}$	$-\frac{2^i}{\ln p}$	**CP**

Fig. 3 Asymptotic data complexity for some statistical attacks.

Explanation of Figure 3

Linear cryptanalysis. In the case of linear cryptanalysis, p_0 is close to $p = 1/2$. If we use the notation of linear cryptanalysis ($p_0 - p = \varepsilon$), we recover $1/2\varepsilon^2$, which is a well-known result due to Matsui [3,4].

Differential cryptanalysis. In this case, both p_0 and p are small but the difference $p_0 - p$ is dominated by p_0. The result we found is slightly different from the standard result, e.g. $1/p_0$ in [8] because it involves $\ln(p_0/p)$. However, the commonly used result requires some restrictions on the ratio p_0/p so it is natural that such a dependency appears.

Differential-linear cryptanalysis. This attack presented in [18] combines a 3-round differential characteristic of probability 1 with a 3-round linear approximation. This case is very similar to linear cryptanalysis since we observe a linear behavior in the output.

Truncated differential cryptanalysis. In the case of truncated differential cryptanalysis, p_0 and p are small but close to each other [9].

Impossible differential. This case is a particular one. The impossible differential cryptanalysis [20] relies on the fact that some event cannot occur in the output of the key dependent permutation. We have always assumed that $p_0 > p$ but in this case it is not true anymore ($p_0 = 0$). However, the formula holds in this case too.

Higher order differential. This attack introduced in [9] is a generalization of differential cryptanalysis. It exploits the fact that a i-th order differential of the cipher is constant (i.e independent from the plaintext and the key). A typical case is when $i = deg(F+1))$, any i-th order differential of F vanishes. Therefore, for this attack, we have $p_0 = 1$. Moreover, $p = (2^m - 1)^{-1}$ where m is the block size so p is small. An important remark here, is that in a cryptanalysis of order i, a sample corresponds to 2^i chosen plaintexts.

5 Success probability of a key-recovery attack

In this section we deal with the key ranking paradigm introduced in Section 2. We present a simple formula that is a good estimate of the success probability expressed in terms of n, the number of key candidates, ℓ, the size of the list to keep and N the number of samples.

5.1 Ordered statistics

Let us denote by $(\xi_i)_{0 \le i < n-1}$ the random variables corresponding to the Σ_{k_i}'s. The analysis phase sorts the Σ_{k_i}'s and keeps the ℓ largest. We denote by ξ_i^* the i-th largest value of the ξ_i's. We are interested in the distribution of ξ_ℓ^* because we keep only a list of keys of size ℓ. The right key k_0 is in the list if and only if $\xi_0 \ge \xi_\ell^*$. The success probability is then:

$$P_S = P\left[\xi_\ell^* \le \xi_0\right] = \sum_{i=0}^{N} P\left[\xi_0 = i\right] \cdot P\left[\xi_\ell^* \le i\right].$$

Let us denote by F the cumulative distribution function of ξ_i's $(i \neq 0)$

$$F(x) = P\left[\xi_1 \le x\right] = \cdots = P\left[\xi_{n-1} \le x\right].$$

It is well known (see for instance [21]) that $F(\xi_\ell^*)$ follows a beta distribution with parameters $n - \ell - 1$ and $\ell - 1$. Let us denote by g this density function. We denote by f_0 the function $f_0(x) = P\left[\xi_0 = \lfloor x \rfloor\right]$. Then, we can write

$$\begin{aligned} P_S &= \sum_{i=0}^{N} f_0(i) \cdot P\left[\xi_\ell^* < i\right] \\ &= \sum_{i=0}^{N} f_0(i) \cdot P\left[F(\xi_\ell^*) < F(i)\right] \\ &= \sum_{i=0}^{N} f_0(i) \cdot \int_0^{F(i)} g(t)\, dt \qquad (12) \end{aligned}$$

5.2 Success probability

The aim of this section is to derive a simple expression giving an estimate of the success probability of a statistical cryptanalysis.

More precisely, we extend a result given by Selçuk in [6] which was a normal distribution approximation of the binomial distribution. To derive the formula of the success probability, it is assumed in [6] that the ℓ-th order statistic is in the limit normally distributed.
Let us denote by $\tilde{f}_0$ the density of the normal distribution with mean Np_0 and variance $Np_0(1-p_0)$. We also define $\tilde{F}^{-1}$ the inverse cumulative normal distribution function with mean Np and variance $Np(1-p)$. Then the work of Selçuk gives the following approximation for the success probability:

$$P_S \approx \int_{\tilde{F}^{-1}(1-\ell/n)}^{\infty} \tilde{f}_0(x)\, dx. \tag{13}$$

Taking the normal distribution as an estimate of the binomial distribution may be misleading for some sets of parameters as stated previously. In this section we derive a similar formula without the help of the Gaussian distribution. Our result is based on the fact that the beta distribution is concentrated around $t_0 \overset{\text{def}}{=} \frac{n-\ell-1}{n-2}$. A last definition is required before giving the principal result of this section.

Definition 4 Let F be the cumulative function of a binomial law with parameters (N,p), that is

$$F(x) \overset{\text{def}}{=} \sum_{i \le x} \binom{N}{i} p^i (1-p)^{N-i}.$$

We define the inverse function F^{-1} by

$$F^{-1}(x) = \min\{t \in \mathbb{N} | F(t) \ge x\}.$$

Remark: It is easy to see that the equality $F(F^{-1}(x)) = x$ may not hold. The definition of F^{-1} implies that $\sum_{i=0}^{F^{-1}(x)} f(x) \ge x$ and $\sum_{i=0}^{F^{-1}(x)-1} f(x) < x$. Hence, we can bound the error term,

$$F(F^{-1}(x)) - x < f(F^{-1}(x)). \tag{14}$$

Theorem 3 *Let P_S be the success probability of a statistical attack that keeps ℓ keys candidates among n. Let N be the number of available samples. We denote by $f_0(i)$ the probability that the key counter corresponding to the good key takes value i, that is $f_0(i) = \binom{N}{i} {p_0}^i (1-p_0)^{N-i}$. We denote by F the cumulative distribution function of the key counters corresponding to the other keys and* *by F^{-1} its inverse function given in Definition 4. Let*

$$\begin{aligned}
\lambda &\overset{\text{def}}{=} \frac{\ell-1}{n-2} = 1-t_0 \\
B &\overset{\text{def}}{=} F^{-1}(1-\lambda) && (15)\\
\delta &\overset{\text{def}}{=} \sum_{i=0}^{B-1} f_0(i) && (16)\\
C_\lambda &\overset{\text{def}}{=} \frac{p}{p}\frac{p_0(N+1)-B}{B-p_0(N+1)} && (17)
\end{aligned}$$

If $\lambda \leq \frac{1}{4}$ then

$$P_S = 1-\delta + O\left(\delta(1+C_\lambda)\sqrt{\frac{\ln(\ell/\delta^2)}{\ell}} + \frac{1}{l^2} + \frac{1}{n}\right).$$

Discussion

On the values taken by C_λ. It turns out that C_λ is for all parameters of cryptographic interest a small constant. To avoid too complicated statements, we avoid giving here general upper-bounds on C_λ. Roughly speaking, this constant is the biggest in the case of linear cryptanalysis, when p_0 and p are very close to each other. In this case, the Gaussian approximation is quite good. If we bring in the Gaussian cumulative function

$$Q(x) \overset{\text{def}}{=} \int_x^\infty \frac{e^{-u^2/2}}{\sqrt{2\pi}} du,$$

then it can be checked from the very definition of B (Equation (15)) that

$$B \approx pN + x\sqrt{Np(1-p)}$$

where $x \overset{\text{def}}{=} Q^{-1}(\lambda)$. Notice that $x \underset{\lambda\to 0^+}{\sim} \sqrt{-2\ln\lambda}$. Moreover from the definition of δ (Equation (16)) we also get that

$$B \approx p_0 N - y\sqrt{Np_0(1-p_0)},$$

where $y \overset{\text{def}}{=} Q^{-1}(\delta)$. We also have $y \underset{\lambda\to 0^+}{\sim} \sqrt{-2\ln\delta}$. Putting all these facts together, we obtain

$$C_\lambda \approx \frac{p}{p_0}\frac{y\sqrt{Np_0(1-p_0)}}{x\sqrt{Np(1-p)}} \approx \sqrt{\frac{-\ln\delta}{-\ln\lambda}}$$

(where we also used that $p_0 \approx p$). Notice that δ can be viewed as an approximation of $1-P_S$ and thus is generally aimed to be around 0.05. For complexity reasons, λ has to be kept small, for instance $\lambda = 10^{-5}$. In this case we have $C_\lambda \approx 0.5$.

Expression of the error term. In [6] an estimate of the success probability is given. In this paper, we give a generalisation of this estimate but we also compute the error $P_S - \sum_{i=F^{-1}(1-\frac{\ell-1}{n-2})}^{N} f_0(i)$. This error term decreases when n and ℓ tend to infinity but it also decreases with δ. Let us recall that $\delta \approx 1 - P_s$ thus, the error induced by using our formula decreases when the success probability grows.

Link with Section 4 In the previous section we express the data complexity in terms of non detection and false alarm error probabilities. Let us recall that β is the probability to accept a wrong key in the list of kept candidates. In this case, the size of the list of kept candidates is not fixed and has a mean of βn. Thus, it seems natural to take $\beta = \ell/n$. Moreover, α is the probability to reject the correct key and thus, α may be chosen to be equal to $1-P_S$. If we use (7) to express α in terms of β, we obtain $\alpha = \sum_{i=0}^{F^{-1}(1-\beta)-1} f_0(i)$. Using the suggested values for both probabilities, this leads to $P_S = 1 - \sum_{i=0}^{F^{-1}(1-\ell/n)-1} f_0(i)$ what corresponds to the result of Theorem 3.

Proof of Theorem 3.
The idea is to split the sum around the critical point t_0. Let $\varepsilon > 0$,

$$\begin{aligned} P_S &= \sum_{i=0}^{N} f_0(i) \int_0^{F(i)} g(t)\, dt \\ &= \sum_{i=0}^{F^{-1}(t_0-\varepsilon)-1} f_0(i) \int_0^{F(i)} g(t)\, dt + \sum_{i=F^{-1}(t_0-\varepsilon)}^{F^{-1}(t_0)-1} f_0(i) \int_0^{F(i)} g(t)\, dt \\ &\quad + \sum_{i=F^{-1}(t_0)}^{N} f_0(i) \int_0^{F(i)} g(t)\, dt \end{aligned}$$

We focus on the third term of the sum

$$\sum_{i=F^{-1}(t_0)}^{N} f_0(i) \int_0^{F(i)} g(t)\, dt = \sum_{i=F^{-1}(t_0)}^{N} f_0(i) - \sum_{i=F^{-1}(t_0)}^{N} f_0(i) \int_{F(i)}^{1} g(t)\, dt.$$

The success probability is essentially $\sum_{i=F^{-1}(t_0)}^{N} f_0(i)$ thus we will now prove that the other terms are negligible.

$$\begin{aligned} P_S - \sum_{i=F^{-1}(t_0)}^{N} f_0(i) = &\underbrace{\sum_{i=0}^{F^{-1}(t_0-\varepsilon)-1} f_0(i) \int_0^{F(i)} g(t)\, dt}_{S_1} \\ &+ \underbrace{\sum_{i=F^{-1}(t_0-\varepsilon)}^{F^{-1}(t_0)-1} f_0(i) \int_0^{F(i)} g(t)\, dt}_{S_2} - \underbrace{\sum_{i=F^{-1}(t_0)}^{N} f_0(i) \int_{F(i)}^{1} g(t)\, dt}_{S_3} \end{aligned}$$

The first argument is that the beta distribution is concentrated around t_0. This means that integrals with domains far enough from t_0 are negligible. This is the case of the integral in S_1, but also for some terms in the sum S_3 (denoted by S_5).

$$S_3 = \underbrace{\sum_{i=F^{-1}(t_0)}^{F^{-1}(t_0+\varepsilon)-1} f_0(i) \int_{F(i)}^{1} g(t)\ dt}_{S_4} + \underbrace{\sum_{i=F^{-1}(t_0+\varepsilon)}^{N} f_0(i) \int_{F(i)}^{1} g(t)\ dt}_{S_5}$$

To sum-up, we now have an error term of $S_1 + S_2 - S_4 - S_5$ with S_1 and S_5 negligible because of the beta distribution. We focus now on $S_2 - S_4$.

$$\begin{aligned}|S_2 - S_4| &\le \max(S_2, S_4) \\ &\le \max\left(\sum_{i=F^{-1}(t_0-\varepsilon)}^{F^{-1}(t_0)-1} f_0(i), \sum_{i=F^{-1}(t_0)}^{F^{-1}(t_0+\varepsilon)-1} f_0(i)\right)\end{aligned}$$

Here the argument is that the sums vanish or are negligible compared to δ. The following lemmas justify the two arguments given in the proof. The first one gives an estimate for the beta distribution tails.

Lemma 5 *Let g be the density function of the beta distribution of parameters $(n-\ell-1, \ell-1)$:*

$$g(t) \overset{\text{def}}{=} (n-1) \cdot \binom{n-2}{\ell-1} \cdot t^{n-\ell-1}(1-t)^{\ell-1}.$$

The maximum of g is reached at $t_0 \overset{\text{def}}{=} \dfrac{n-\ell-1}{n-2}$. Let $\varepsilon \overset{\text{def}}{=} z \cdot \dfrac{\sqrt{\ell-1}}{n-2}$. If $z = o\left(\sqrt{\ell}\right)$ and $\ell \in [1, n/2]$, we have:

$$\int_{t_0-\varepsilon}^{t_0+\varepsilon} g(t)\ dt = 1 + O\left(\frac{1}{\ell^2} + \frac{1}{n} + \frac{e^{-z^2/2}}{z}\right).$$

Proof see Appendix 7.5. □

The second one expresses S_2 as a function of $\delta \approx 1 - P_S$. Notice that this can be done for S_4 in a similar way.

Lemma 6 *Let $\varepsilon = z\frac{\sqrt{\ell-1}}{n}$ for some value z where $z = o(\sqrt{\ell})$ when ℓ goes to infinity. If $\lambda \le \frac{1}{4}$, then*

$$\sum_{i=F^{-1}(t_0-\varepsilon)}^{F^{-1}(t_0)-1} f_0(i) = O\left(\frac{zC_\lambda\delta}{\sqrt{\ell-1}}\right).$$

Proof see Appendix 7.6. □

We go back to the proof of Theorem 3. Let us recall that we want to bound the error

$$P_S - \sum_{i=F^{-1}(t_0)}^{N} f_0(i) = S_1 + S_2 - S_4 - S_5.$$

We bound S_1 and S_5:

$$\begin{aligned} S_1 &= \sum_{i=0}^{F^{-1}(t_0-\varepsilon)-1} f_0(i) \int_0^{F(i)} g(t)\ dt \leq \int_0^{t_0-\varepsilon} g(t)\ dt, \\ S_5 &= \sum_{i=F^{-1}(t_0+\varepsilon)}^{N} f_0(i) \int_{F(i)}^{1} g(t)\ dt \leq \int_{t_0+\varepsilon}^{1} g(t)\ dt. \end{aligned}$$

Moreover, $|S_1 - S_5| \leq S_1 + S_5 \leq 1 - \int_{t_0-\varepsilon}^{t_0+\varepsilon} g(t)\ dt$. Thus, using Lemma 5,

$$|S_1 - S_5| = O\left(\frac{1}{\ell^2} + \frac{1}{n} + \frac{e^{-z^2/2}}{z}\right) \tag{18}$$

To show that S_2 is negligible we use Lemma 6. This lemma can be slightly modified to prove that S_4 is negligible too. Hence,

$$|S_2 - S_4| = O\left(\delta \frac{z}{\sqrt{\ell}}\right). \tag{19}$$

Adding (18) and (19) gives the following result

$$P_S - \sum_{i=F^{-1}(t_0)}^{N} f_0(i) = O\left(\frac{1}{\ell^2} + \frac{1}{n} + \frac{e^{-z^2/2}}{z} + \delta \frac{z}{\sqrt{\ell}}\right).$$

The final step consists in choosing a particular z. Taking $z = \sqrt{\ln\left(\frac{\ell}{\delta^2}\right)}$ gives [1]

$$P_S - \sum_{i=F^{-1}(t_0)}^{N} f_0(i) = O\left(\delta\sqrt{\frac{\ln(\ell/\delta^2)}{\ell}} + \frac{1}{\ell^2} + \frac{1}{n}\right).$$

□

[1] The point of choosing z like this is that it can be easily checked that $\frac{e^{-z^2/2}}{z} = O\left(\delta \frac{z}{\sqrt{\ell}}\right)$.

5.3 Experimental results

In Figure 4 we compare our formula for the success probability

$$P_s \approx \sum_{i=F^{-1}(1-\frac{\ell-1}{n-2})}^{N} f_0(i) \tag{20}$$

with Formula (13) given by Selçuk [6] and the true value. This value is numerically computed using (12) with large precision.
In the case of linear cryptanalysis, as the Gaussian approximation is valid, our expression of the success probability is the same as the one given by Selçuk. However, in the case of differential cryptanalysis, the formula given by Selçuk is too optimistic, while our Expression (20) is close to the true value.

Type of cryptanalysis	Probabilities	Parameters $N = 2^{48}$ $n = 2^{20}$	P_S	our estimate of P_S (20)	estimate [6] of P_S (13)
Linear	$p = 0.5$ $p_0 = p + 1.49 \cdot 2^{-24}$	$\ell = 2^{15}$	0.8681	0.8681	0.8681
Linear	$p = 0.5$ $p_0 = p + 1.49 \cdot 2^{-24}$	$\ell = 2^{10}$	0.4533	0.4533	0.4533
Differential	$p = 2^{-64}$ $p_0 = 2^{-47.2}$	$\ell = 2^{15}$	0.8257	0.8247	0.9050
Differential	$p = 2^{-64}$ $p_0 = 2^{-47.2}$	$\ell = 2^{10}$	0.8250	0.8247	0.9050

Fig. 4 Comparision of the estimates (20) and (13) with the true value of the success probability.

The point of Figure 5 is to demonstrate that when we choose N of the form [2]

$$N = -c \cdot \frac{\ln(2\sqrt{\pi} \cdot \ell/n)}{D\left(p_0||p\right)},$$

then the success probability P_S depends essentially only on c and is basically independent of the type of cryptanalysis. We have computed in Figure 5 several values of the success probability for $n = 2^{60}$ for various values of ℓ and types of cryptanalysis.

6 Conclusion

In this paper, we give a general framework to estimate the number of samples that are required to perform a statistical cryptanalysis. We use this framework to provide a simple algorithm which accurately computes the number of

[2] This choice is guided by Theorem 2 and (11) where we have shown that data complexity is of order $O\left(-\frac{\ln(2\sqrt{\pi}\cdot\ell/n)}{D(p_0||p)}\right)$.

	$c = 1$			$c = 1.5$		
Parameters		ℓ			ℓ	
	2^{10}	2^{20}	2^{30}	2^{10}	2^{20}	2^{30}
$p = 0.5$ $p_0 = p + 1.49 \cdot 2^{-24}$	0.5855	0.5898	0.5949	0.9799	0.9687	0.9500
$p = 0.5$ $p_0 = p + 1.23 \cdot 2^{-11}$	0.5856	0.5899	0.5950	0.9799	0.9687	0.9500
$p = 2^{-30}$ $p_0 = 1.2 \cdot 2^{-30}$	0.5802	0.5921	0.5875	0.9766	0.9650	0.9446
$p = 2^{-40}$ $p_0 = 1.2 \cdot 2^{-40}$	0.5802	0.5827	0.5875	0.9766	0.9650	0.9446
$p = 2^{-64}$ $p_0 = 2^{-60}$	0.5801	0.5257	0.5544	0.9249	0.9070	0.8844
$p = 2^{-32}$ $p_0 = 2^{-29}$	0.5993	0.6179	0.6443	0.9605	0.9375	0.9241

Fig. 5 Success probability for various parameters with $n = 2^{60}$ and $N = -c \cdot \frac{\ln(2\sqrt{\pi} \cdot \ell/n)}{D(p_0||p)}$

samples which is required for achieving some given error probabilities. Furthermore, we provide an explicit formula (Theorem 2) which gives a good estimate of the number of required samples (bounds on relative error are given). A further simplification of the data complexity shows that the behavior of the number of samples is dominated by $D\left(p_0||p\right)^{-1}$. We show that $D\left(p_0||p\right)^{-1}$ gives the same order of magnitude as known results excepted in differential cryptanalysis where a dependency on $\ln(p_0/p)$ is emphasized. We also extend these results to other statistical cryptanalyses, for instance, truncated differential cryptanalysis.

On the other hand, we provide a simple formula for the success probability in terms of n, the number of key candidates, ℓ, the size of the list of kept candidates, and N, the number of available samples (Theorem 3). This formula is a generalization of a formula obtained by Selçuk using in particular a Gaussian approximation for the binomial distribution. Since we do not use such an approximation, our result is valid for all sets of parameters (p_0, p), including differential cryptanalysis. Moreover, we give an estimate of the error made using this formula for the success probability. Finally, using this expression, we compute some success probabilities for some sets of parameters and notice that when N is of the form

$$N = c \cdot \frac{-\ln(2\sqrt{\pi} l/n)}{D\left(p_0||p\right)}$$

as suggested by Theorem 2 and (11), the success probability seems to only depend on c.

References

1. Vaudenay, S.: Decorrelation: A Theory for Block Cipher Security. Journal of Cryptology **16** (2003) 249–286
2. Tardy-Corfdir, A., Gilbert, H.: A Known Plaintext Attack of FEAL-4 and FEAL-6. In: CRYPTO '91. Volume 576 of LNCS., Springer–Verlag (1992) 172–181
3. Matsui, M.: Linear cryptanalysis method for DES cipher. In: EUROCRYPT '93. Volume 765 of LNCS., Springer–Verlag (1993) 386–397
4. Matsui, M.: The First Experimental Cryptanalysis of the Data Encryption Standard. In: CRYPTO '94. Volume 839 of LNCS., Springer–Verlag (1994) 1–11
5. Biham, E., Shamir, A.: Differential Cryptanalysis of DES-like Cryptosystems. Journal of Cryptology **4** (1991) 3–72
6. Selçuk, A.A.: On Probability of Success in Linear and Differential Cryptanalysis. Journal of Cryptology **21** (2008) 131–147
7. Harpes, C., Kramer, G., Massey, J.: A Generalization of Linear Cryptanalysis and the Applicability of Matsui's Piling-up Lemma. In: EUROCRYPT '95. Volume 921 of LNCS., Springer–Verlag (1995) 24–38
8. Lai, X., Massey, J.L., Murphy, S.: Markov Ciphers and Differential Cryptanalysis. LNCS **547** (1991) 17–38
9. Knudsen, L.R.: Truncated and Higher Order Differentials. In: FSE '94. Volume 1008 of LNCS., Springer–Verlag (1994) 196–211
10. Junod, P.: On the Optimality of Linear, Differential, and Sequential Distinguishers. In: EUROCRYPT '03. Volume 2656 of LNCS., Springer–Verlag (2003) 17–32
11. Baignères, T., Junod, P., Vaudenay, S.: How Far Can We Go Beyond Linear Cryptanalysis? In: ASIACRYPT '04. Volume 3329 of LNCS., Springer–Verlag (2004) 432–450
12. Baignères, T., Vaudenay, S.: The Complexity of Distinguishing Distributions. In: ICITS '08. Volume 5155 of LNCS., SV (2008) 210–222
13. Junod, P.: On the Complexity of Matsui's Attack. In: SAC '01. Volume 2259 of LNCS., Springer–Verlag (2001) 199–211
14. Junod, P., Vaudenay, S.: Optimal key ranking procedures in a statistical cryptanalysis. In: FSE '03. Volume 2887 of LNCS., Springer–Verlag (2003) 235–246
15. Nyberg, K.: Generalized Feistel Networks. In: ASIACRYPT '96. Volume 1163 of LNCS., Springer–Verlag (1996) 91–104
16. Cover, T., Thomas, J.: Information theory. Wiley series in communications. Wiley (1991)
17. Arriata, R., Gordon, L.: Tutorial on large deviations for the binomial distribution. Bulletin of Mathematical Biology **51** (1989) 125–131
18. Langford, S.K., Hellman, M.E.: Differential-Linear Cryptanalysis. In: CRYPTO '94. Volume 839 of LNCS., Springer–Verlag (1994) 17–25
19. Biham, E., Shamir, A.: Differential Cryptanalysis of the Full 16-round DES. In: CRYPTO'92. Volume 740 of LNCS., Springer–Verlag (1993) 487–496
20. Biham, E., Biryukov, A., Shamir, A.: Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differentials. In: EUROCRYPT '99. Volume 1592 of LNCS. (1999) 12–23
21. David, H., Nagaraja, H.: Order Statistics (third edition). Wiley series in Probability Theory. John Wiley and Sons (2003)
22. Gilbert, H.: Cryptanalyse statistique des algorithmes de chiffrement et sécurité des schémas d'authentification. PhD thesis, Université Paris 11 Orsay (1997)
23. Biham, E., Dunkelman, O., Keller, N.: Enhancing Differential-Linear Cryptanalysis. In: ASIACRYPT '02. Volume 2501 of LNCS., SV (2002) 254–266
24. Junod, P.: Statistical cryptanalysis of block ciphers. PhD thesis, EPFL (2005)

7 Appendix

7.1 Generalized Feistel networks

A generalized Feistel network [15] is an iterated block cipher whose round function is depicted in Figure 6.

Definition 5 In a generalized Feistel network with block size $2dn$, the plaintext X is split into $2n$ blocks of size d. It uses n S-boxes of dimension $d \times d$ denoted by $S_1, ..., S_n$ and the round function $(X_1, ..., X_{2n}) \mapsto (Y_1, ..., Y_{2n})$ is defined by:

$$\begin{aligned} Z_{n+1-i} &= X_{n+1-i} \oplus S_i(X_{i+n} \oplus K_i) \textbf{ for } i = 1, ..., n \\ Z_i &= X_i \textbf{ for } i = n+1, ..., 2n \\ Y_i &= Z_{i-1} \textbf{ for } i \neq 1 \\ Y_1 &= Z_{2n} \end{aligned}$$

where $\oplus$ is the modulo 2 addition.

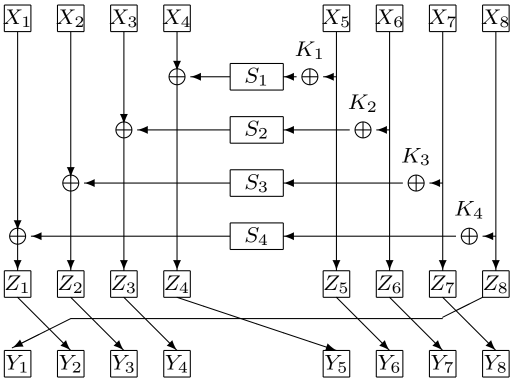

Fig. 6 Generalized Feistel network with 4 S-boxes

7.2 Discussion on Algorithm 1: Finding N_{nd} and N_{fa}

A more efficient technique than dichotomic search can be used to find N_{nd} and N_{fa} in Algorithm 1. If we fix the non-detection error probability to α, (4) can be rewritten as:

$$N \sim \frac{1}{D(\tau||p_0)} \ln\left(\frac{p_0\sqrt{1-\tau}}{\alpha(p_0-\tau)\sqrt{2\pi N\tau}}\right)$$

Using the same fixed point argument as in Appendix 7.4, we can find N_{nd} by iterating the function with a first point $x_0 = D(\tau||p_0)^{-1}$. The same thing can be done with (3) in order to find N_{fa}.

7.3 Taylor expansions of the Kullback-Leibler divergence

This appendix contains three lemmas giving the asymptotic behavior of some expressions that involve the Kullback-Leibler divergence and that are used all over the paper.

Lemma 7 *Let $0 < a < b < 1$ such that $O\left(\frac{b-a}{1-a}\right) = O\left(b-a\right)$. Then,*

$$D\left(b||a\right) = b\left[\ln\left(\frac{b}{a}\right) - \frac{b-a}{b} + \frac{(b-a)^2}{2b(1-b)}\right] + O(b-a)^3$$

Proof Using the Taylor theorem, we get

$$(1-b)\ln\left(\frac{1-b}{1-a}\right) = a - b + \frac{(a-b)^2}{2(1-b)} + O(b-a)^3.$$

Therefore,

$$\begin{aligned} D\left(b||a\right) &= b\ln\left(\frac{b}{a}\right) + (1-b)\ln\left(\frac{1-b}{1-a}\right) \\ &= b\ln\left(\frac{b}{a}\right) + a - b + \frac{(a-b)^2}{2(1-b)} + O\left(b-a\right)^3 \\ &= b\left[\ln\left(\frac{b}{a}\right) - \frac{b-a}{b} + \frac{(b-a)^2}{2b(1-b)} + O(b-a)^3\right]. \end{aligned}$$

Lemma 8 *Let $\varepsilon > 0$ be a real number such that $O\left(\frac{\varepsilon}{a}\right) = O\left(\frac{\varepsilon}{1-a}\right) = O\left(\varepsilon\right)$. Then,*

$$D\left(a+\varepsilon||a\right) = \frac{\varepsilon^2}{2a(1-a)} + O\left(\varepsilon^3\right).$$

Proof Using Lemma 7, we have that

$$D\left(a+\varepsilon||a\right) = (a+\varepsilon)\left[\ln\left(1+\frac{\varepsilon}{a}\right) - \frac{\varepsilon}{a+\varepsilon} + \frac{\varepsilon^2}{2(a+\varepsilon)(1-a-\varepsilon)}\right] + O\left(\varepsilon^3\right).$$

Since $\varepsilon/a = O\left(\varepsilon\right)$, we expand the logarithm to get

$$\begin{aligned} D\left(a+\varepsilon||a\right) &= (a+\varepsilon)\left[\frac{\varepsilon}{a} - \frac{\varepsilon^2}{2a^2} - \frac{\varepsilon}{a+\varepsilon} + \frac{\varepsilon^2}{2(a+\varepsilon)(1-a-\varepsilon)} + O\left(\frac{\varepsilon^3}{a^3}\right)\right] + O\left(\varepsilon^3\right) \\ &= (a+\varepsilon)\left[\frac{\varepsilon^2\ a}{2a^2(a+\varepsilon)(1-a-\varepsilon)} + O\left(\varepsilon^3\right)\right] + O\left(\varepsilon^3\right) \\ &= \frac{\varepsilon^2}{2a(1-a)} + O\left(\varepsilon^3\right). \end{aligned}$$

□

Lemma 9 *If* $O\left(\frac{\varepsilon}{a}\right) = O\left(\frac{\varepsilon}{1-a}\right) = O\left(\varepsilon\right)$*, then,*

$$\varDelta_\varepsilon = D\left(a||a-\varepsilon\right) - D\left(a||a+\varepsilon\right) = \frac{2}{3}\varepsilon^3 \cdot \frac{1-2a}{a^2(1-a)^2} + O\left(\varepsilon^4\right).$$

Proof We split $\varDelta_\varepsilon$ into two terms.

$$\begin{aligned}
\varDelta_\varepsilon &= D\left(a||a-\varepsilon\right) - D\left(a||a+\varepsilon\right) \\
&= a\ln\left(\frac{a+\varepsilon}{a-\varepsilon}\right) + (1-a)\ln\left(\frac{1-a-\varepsilon}{1-a+\varepsilon}\right) \\
&= \varDelta_{\varepsilon,a} + \varDelta_{\varepsilon,1-a}
\end{aligned}$$

Expanding the logarithm in the first term gives:

$$\begin{aligned}
\varDelta_{\varepsilon,a} &= a\ln\left[1 + \frac{2\varepsilon}{a-\varepsilon}\right] \\
&= a\left[\frac{2\varepsilon}{a-\varepsilon} - \frac{2\varepsilon^2}{(a-\varepsilon)^2} + \frac{8\varepsilon^3}{3(a-\varepsilon)^3} + O\left(\varepsilon^4\right)\right] \\
&= \frac{1}{1-\frac{\varepsilon}{a}}\left[2\varepsilon - \frac{2\varepsilon^2}{a-\varepsilon} + \frac{8\varepsilon^3}{3(a-\varepsilon)^2} + O\left(\varepsilon^4\right)\right] \\
&= \left[1 + \frac{\varepsilon}{a} + \frac{\varepsilon^2}{a^2} + o\left(\varepsilon^2\right)\right] \cdot \left[2\varepsilon - \frac{2\varepsilon^2}{a-\varepsilon} + \frac{8\varepsilon^3}{3(a-\varepsilon)^2} + O\left(\varepsilon^4\right)\right] \\
&= 2\varepsilon - \frac{2\varepsilon^2}{(a-\varepsilon)} + \frac{8\varepsilon^3}{3(a-\varepsilon)^2} + \frac{2\varepsilon^2}{a} - \frac{2\varepsilon^3}{a(a-\varepsilon)} + \frac{2\varepsilon^3}{a^2} + O\left(\varepsilon^4\right) \\
&= 2\varepsilon\left[1 + \varepsilon\left(\frac{1}{a} - \frac{1}{a-\varepsilon}\right) + \varepsilon^2\frac{4}{3a^2} + O\left(\varepsilon^3\right)\right].
\end{aligned}$$

Similarly, we get

$$\varDelta_{\varepsilon,1-a} = 2\varepsilon\left[-1 + \varepsilon\left(\frac{1}{1-a} - \frac{1}{1-a+\varepsilon}\right) - \varepsilon^2\frac{4}{3(1-a)^2} + O\left(\varepsilon^3\right)\right]$$

Summing the two terms we obtain,

$$\begin{aligned}
\varDelta_\varepsilon &= \varDelta_{\varepsilon,a} + \varDelta_{\varepsilon,1-a} \\
&= 2\varepsilon^2\left[\frac{1}{a} - \frac{1}{a-\varepsilon} + \frac{1}{1-a} - \frac{1}{1-a+\varepsilon} + \varepsilon\left(\frac{4}{3a^2} - \frac{4}{3(1-a)^2}\right) + O\left(\varepsilon^2\right)\right] \\
&= 2\varepsilon^2\left[\varepsilon \cdot \frac{2a-1}{a^2(1-a)^2} + \varepsilon\left(\frac{4}{3a^2} - \frac{4}{3(1-a)^2}\right) + O\left(\varepsilon^2\right)\right].
\end{aligned}$$

And finally,

$$\varDelta_\varepsilon = \frac{2}{3}\varepsilon^3 \cdot \frac{1-2a}{a^2(1-a)^2} + O\left(\varepsilon^4\right).$$

□

7.4 Proof of Theorem 2

Proof Recall that $\tau = p_0$ so that non-detection error probability is around $\frac{1}{2}$. We want to control false alarm error probability that we fix to β. Equation (3) in Theorem 1 gives

$$N \approx -\frac{\ln(\nu\beta\sqrt{N})}{D\left(p_0||p\right)} \tag{21}$$

where $\nu \overset{\text{def}}{=} \frac{(p_0-p)\sqrt{2\pi(1-p_0)}}{(1-p)\sqrt{p_0}}$. Formula (21) suggests to bring in the contractive function f:

$$f(x) \overset{\text{def}}{=} -\frac{\ln(\nu\beta\sqrt{x})}{D\left(p_0||p\right)}.$$

Applying f iteratively with first term $N_0 = 1$ gives a sequence $(N_i)_{i\geq 0}$ which can be shown to have a limit N_∞ which is the required number of samples. Since f is decreasing, consecutive terms satisfy $N_{2i} \leq N_\infty \leq N_{2i+1}$. Function f can be written as

$$f(x) = a - b\ln(x) \text{ with } a \overset{\text{def}}{=} -\frac{\ln(\nu\beta)}{D\left(p_0||p\right)} \text{ and } b \overset{\text{def}}{=} \frac{1}{2D\left(p_0||p\right)}.$$

It is worth noticing that a corresponds to the second term, N_1, of the sequence. Now, we want to show that the third term, N_2, provides a good approximation of N_∞. As $N_2 \leq N_\infty \leq N_3$, it is desirable to express N_3 in terms of N_2.

$$\begin{aligned} N_3 &= N_1 - b\ln(N_1) + b\ln\left(N_1/N_2\right) \\ &= N_2 + b\ln\left(N_1/N_2\right) \end{aligned}$$

Let us define $\theta = \left[1 + \frac{1}{2\ln(\nu\beta)}\ln\left(-\frac{\ln(\nu\beta)}{D\left(p_0||p\right)}\right)\right]^{-1}$, as in Equation (10) in Theorem 2. Then,

$$\begin{aligned} \frac{N_2}{N_1} &= 1 + \frac{b\ln(a)}{a} = 1 + \frac{\ln(a)}{2\ln(\nu\beta)} \\ &= \left[1 + \frac{1}{2\ln(\nu\beta)}\ln\left(-\frac{\ln(\nu\beta)}{D\left(p_0||p\right)}\right)\right] = \theta^{-1}. \end{aligned}$$

The bound on N_∞ becomes:

$$N_2 \leq N_\infty \leq N_2\left[1 + \frac{b\ln(\theta)}{N_2}\right].$$

in order to show that N_2 is a good approximation of N_∞, we focus on $b\ln(\theta)/N_2$ and compare it with 1. Since $N_2/b = a/b - \ln(a)$, we try to bound a/b. We have $\theta N_2 = N_1$ implying $a/b = \theta\ln(a)/(\theta - 1)$. Since f is a decreasing function, $N_1 > N_2$ leading to $N_2/b \geq \ln(N_2)/(\theta - 1)$.

Finally, $N_3 \leq N_2 \left[1 + \frac{(\theta-1)\ln(\theta)}{\ln(N_2)}\right]$ and

$$N_2 \leq N_\infty \leq N_2 \left[1 + \frac{(\theta-1)\ln(\theta)}{\ln(N_2)}\right]$$

where N_2 is equal to the value of N' in Theorem 2.

7.5 Concentration of the beta density function (proof of Lemma 5)

The proof of Lemma 5 relies heavily on the following expansion.

Lemma 10 *Let $\phi(t)$ be a function defined on $(0,1)$ that is four times differentiable. Suppose that this function has a minimum value of 0 reached at $t_0 \in (\frac{1}{2}, 1)$ and that $\phi''(t_0) > 0$. Let λ be a positive real number. Then, for $\varepsilon \in (0, 1-t_0)$,*

$$\int_{t_0}^{t_0+\varepsilon} e^{-\lambda\phi(t)}\, dt = \int_0^{\phi(t_0+\varepsilon)} \left[\frac{1}{\sqrt{2\tau\phi''(t_0)}} - \frac{1}{3}\frac{\phi'''(t_0)}{\phi''^2(t_0)} + A_{t_0}\sqrt{\tau} + o\left(\sqrt{\tau}\right)\right] e^{-\lambda\tau} d\tau$$

and

$$\int_{t_0-\varepsilon}^{t_0} e^{-\lambda\phi(t)}\, dt = \int_0^{\phi(t_0-\varepsilon)} \left[\frac{1}{\sqrt{2\tau\phi''(t_0)}} + \frac{1}{3}\frac{\phi'''(t_0)}{\phi''^2(t_0)} + A_{t_0}\sqrt{\tau} + o\left(\sqrt{\tau}\right)\right] e^{-\lambda\tau} d\tau.$$

Where $A_{t_0} \overset{\text{def}}{=} \frac{\sqrt{2}}{24\phi''(t_0)^{5/2}} \left(\frac{5\phi^{(3)}(t_0)^2}{\phi''(t_0)} - 3\phi^{(4)}(t_0)\right).$

Proof Substituting τ for $\phi(t)$, we obtain

$$\int_{t_0}^{t_0\pm\varepsilon} e^{-\lambda\phi(t)} dt = \int_0^{\phi(t_0\pm\varepsilon)} G(\tau) e^{-\lambda\tau} d\tau$$

with $G(\tau) = \left.\frac{1}{\phi'(t)}\right|_{t=\phi^{-1}(\tau)}$.

First of all, we are going to express $t-t_0$ as a function of τ using the following expansion of ϕ.

$$\phi(t) = \frac{\phi''(t_0)}{2}(t-t_0)^2 + \frac{\phi^{(3)}(t_0)}{6}(t-t_0)^3 + \frac{\phi^{(4)}(t_0)}{24}(t-t_0)^4 + o\left((t-t_0)^4\right).$$

We turn now to the asymptotic behavior of $t-t_0$. Without loss of generality, we assume that $t > t_0$.

$$(t-t_0)^2 = \frac{2\phi(t)}{\phi''(t_0)} \left[1 + \frac{1}{3}\frac{\phi^{(3)}(t_0)}{\phi''(t_0)}(t-t_0) + \frac{1}{12}\frac{\phi^{(4)}(t_0)}{\phi''(t_0)}(t-t_0)^2 + o\left((t-t_0)^2\right)\right]^{-1} \tag{22}$$

This gives $t - t_0 = \sqrt{\dfrac{2\tau}{\phi''(t_0)}}\left[1 + O\left(\sqrt{\tau}\right)\right]$. Plugging this expression back into (22) leads to

$$t - t_0 = \sqrt{\frac{2\tau}{\phi''(t_0)}}\left[1 - \frac{\sqrt{2}}{6}\frac{\phi^{(3)}(t_0)}{\phi''(t_0)^{3/2}}\sqrt{\tau} + o(\sqrt{\tau})\right].$$

Then, going one step further leads to:

$$(t - t_0)^2 = \frac{2\tau}{\phi''(t_0)}\left[1 + \frac{\sqrt{2}}{3}\frac{\phi^{(3)}(t_0)}{\phi''(t_0)^{3/2}}\left[1 - \frac{\sqrt{2}}{6}\frac{\phi^{(3)}(t_0)}{\phi''(t_0)^{3/2}}\sqrt{\tau}\right]\sqrt{\tau} + \frac{1}{6}\frac{\phi^{(4)}(t_0)}{\phi''(t_0)^2}\tau + o\left(\tau\right)\right]^{-1}$$

$$t - t_0 = \sqrt{\frac{2\tau}{\phi''(t_0)}}\left[1 + \frac{\sqrt{2}}{3}\frac{\phi^{(3)}(t_0)}{\phi''(t_0)^{3/2}}\sqrt{\tau} + \left(\frac{1}{6}\frac{\phi^{(4)}(t_0)}{\phi''(t_0)^2} - \frac{1}{9}\frac{\phi^{(3)}(t_0)^2}{\phi''(t_0)^3}\right)\tau + o\left(\tau\right)\right]^{-1/2}$$

And we finally get:

$$t - t_0 = \sqrt{\frac{2\tau}{\phi''(t_0)}}\left[1 - \frac{\sqrt{2}}{6}\frac{\phi^{(3)}(t_0)}{\phi''(t_0)^{3/2}}\sqrt{\tau} - \left(\frac{1}{12}\frac{\phi^{(4)}(t_0)}{\phi''(t_0)^2} - \frac{5}{36}\frac{\phi^{(3)}(t_0)^2}{\phi''(t_0)^3}\right)\tau + o\left(\tau\right)\right]. \tag{23}$$

Using the same method we show that, if $t < t_0$, then

$$t - t_0 = -\sqrt{\frac{2\tau}{\phi''(t_0)}}\left[1 + \frac{\sqrt{2}}{6}\frac{\phi^{(3)}(t_0)}{\phi''(t_0)^2}\sqrt{\tau} - \left(\frac{1}{12}\frac{\phi^{(4)}(t_0)}{\phi''(t_0)^2} - \frac{5}{36}\frac{\phi^{(3)}(t_0)^2}{\phi''(t_0)^3}\right)\tau + o(\tau)\right]. \tag{24}$$

Now that we have an expression of $t - t_0$ as a function of τ we go back to the computation of $G(\tau)$. We use the following Taylor series:

$$\phi'(t) = \phi''(t_0)(t - t_0) + \frac{\phi^{(3)}(t_0)}{2}(t - t_0)^2 + \frac{\phi^{(4)}(t_0)}{6}(t - t_0)^3 + o\left((t - t_0)^3\right).$$

This gives the following expression of $\frac{1}{\phi'(t)}$:

$$\begin{aligned}\frac{1}{\phi'(t)} &= \frac{1}{\phi''(t_0)(t - t_0) + \frac{1}{2}\phi^{(3)}(t_0)(t - t_0)^2 + \frac{1}{6}\phi^{(4)}(t_0)(t - t_0)^3 + o\left((t - t_0)^3\right)} \\ &= \frac{1}{\phi''(t_0)(t - t_0)}\left[1 + \frac{1}{2}\frac{\phi^{(3)}(t_0)}{\phi''(t_0)}(t - t_0) + \frac{1}{6}\frac{\phi^{(4)}(t_0)}{\phi''(t_0)}(t - t_0)^2 + o\left((t - t_0)^2\right)\right]^{-1} \\ &= \frac{1}{\phi''(t_0)(t - t_0)}\left[1 - \frac{\phi^{(3)}(t_0)}{2\phi''(t_0)}(t - t_0) + \left(3\frac{\phi^{(3)}(t_0)^2}{\phi''(t_0)} - 2\phi^{(4)}(t_0)\right)\frac{(t - t_0)^2}{12\phi''(t_0)} + o\left((t - t_0)^2\right)\right] \\ &= \frac{1}{\phi''(t_0)(t - t_0)} - \frac{\phi^{(3)}(t_0)}{2\phi''(t_0)^2} + \left(3\frac{\phi^{(3)}(t_0)^2}{\phi''(t_0)} - 2\phi^{(4)}(t_0)\right)\frac{t - t_0}{12\phi''(t_0)^2} + o\left(t - t_0\right).\end{aligned}$$

We are now going to plug (23) in this formula. The first term can be written as:

$$\frac{1}{\phi''(t_0)(t-t_0)} = \frac{1}{\sqrt{2\phi''(t)\tau}}\left[1-\frac{\sqrt{2}}{6}\frac{\phi^{(3)}(t_0)}{\phi''(t_0)^{3/2}}\sqrt{\tau}-\left(\frac{1}{12}\frac{\phi^{(4)}(t_0)}{\phi''(t_0)^2}-\frac{5}{36}\frac{\phi^{(3)}(t_0)^2}{\phi''(t_0)^3}\right)\tau+o\left(\tau\right)\right]^{-1}$$

$$= \frac{1}{\sqrt{2\phi''(t)\tau}}+\frac{\phi^{(3)}(t_0)}{6\phi''(t_0)^2}+\left(\phi^{(4)}(t_0)-\frac{\phi^{(3)}(t_0)^2}{\phi''(t_0)}\right)\frac{\sqrt{2}}{24\phi''(t_0)^{5/2}}\sqrt{\tau}+o\left(\sqrt{\tau}\right).$$

And the second one:

$$\left(3\frac{\phi^{(3)}(t_0)^2}{\phi''(t_0)}-2\phi^{(4)}(t_0)\right)\frac{t-t_0}{12\phi''(t_0)^2} = \left(6\frac{\phi^{(3)}(t_0)^2}{\phi''(t_0)}-4\phi^{(4)}(t_0)\right)\frac{\sqrt{2}}{24\phi''(t_0)^{5/2}}\sqrt{\tau}+o\left(\tau\right).$$

Putting these results together leads to

$$G(\tau) = \frac{1}{\sqrt{2\phi''(t_0)\tau}}-\frac{\phi^{(3)}(t_0)}{3\phi''(t_0)^2}+\frac{\sqrt{2}}{24\phi''(t_0)^{5/2}}\left(5\frac{\phi^{(3)}(t_0)^2}{\phi''(t_0)}-3\phi^{(4)}(t_0)\right)\sqrt{\tau}+o\left(\tau\right).$$

In the case $t < t_0$, using (24), we get

$$-G(\tau) = \frac{1}{\sqrt{2\phi''(t_0)\tau}}+\frac{\phi^{(3)}(t_0)}{3\phi''(t_0)^2}+\frac{\sqrt{2}}{24\phi''(t_0)^{5/2}}\left(5\frac{\phi^{x}(3)(t_0)^2}{\phi''(t_0)}-3\phi^{(4)}(t_0)\right)\sqrt{\tau}+o\left(\tau\right).$$

□

We are now going to use this result to prove Lemma 5.
Lemma 5

$$g(t) = (n-1)\cdot\binom{n-2}{\ell-1}\cdot t^{n-\ell-1}(1-t)^{\ell-1}, \quad t_0 \overset{\text{def}}{=} 1-\lambda = \frac{n-\ell-1}{n-2}, \text{ and } \varepsilon = z\cdot\frac{\sqrt{\ell-1}}{n-2}.$$

Then, under the conditions $1 \le \ell \le n/2$ and $z > 0, z = o\left(\sqrt{\ell}\right)$, we have

$$\int_{t_0-\varepsilon}^{t_0+\varepsilon} g(t)\,dt = 1 + O\left(\frac{1}{\ell^2}+\frac{1}{n}+\frac{e^{-z^2/2}}{z}\right)$$

Proof First, we apply Stirling approximation to the binomial coefficient.

$$\binom{n-2}{\ell-1} = \sqrt{\frac{1}{2\pi}}\left(\frac{n-2}{n-\ell-1}\right)^{n-\ell-1/2}\left(\frac{n-2}{\ell-1}\right)^{\ell-1/2}\left[1-\frac{1}{12(\ell-1)}+O\left(\frac{1}{n}+\frac{1}{\ell^2}\right)\right].$$

We simplify the expression

$$\left(\frac{n-2}{n-\ell-1}\right)^{n-\ell-1}\left(\frac{n-2}{\ell-1}\right)^{\ell-1}t^{n-\ell-1}(1-t)^{\ell-1} = e^{-(n-2)D(t_0||t)}.$$

This leads us to define a new function $\tilde{g}$.

$$\tilde{g}(t) = C_{n,\ell}\cdot e^{-(n-2)D(t_0||t)}.$$

with $C_{n,\ell} = (n-1) \cdot \sqrt{\frac{n-2}{2\pi(\ell-1)(n-\ell-1)}}$.
Then,

$$g(t) = \tilde{g}(t) \cdot \left[1 - \frac{1}{12(\ell-1)} + O\left(\frac{1}{n} + \frac{1}{\ell^2}\right)\right].$$

The structure of $\tilde{g}$ suggests to use Lemma 10 with $\lambda = n-2$ and $\phi(t) = D(t_0||t)$. Then,

$$\begin{aligned} \phi''(t_0) &= \frac{1}{t_0} + \frac{1}{1-t_0} = \frac{1}{t_0(1-t_0)} > 0, \\ \phi^{(3)}(t_0) &= \frac{2}{(1-t_0)^2} - \frac{2}{t_0^2} = 2\frac{2t_0-1}{t_0^2(1-t_0)^2}, \\ \phi^{(4)}(t_0) &= \frac{6}{(1-t_0)^3} + \frac{6}{t_0^3} = 6\frac{3t_0^2-3t_0+1}{t_0^3(1-t_0)^3}, \\ \text{and } A_{t_0} &= \frac{13t_0^2-13t_0+1}{6\sqrt{2t_0(1-t_0)}}. \end{aligned}$$

Since $\phi''(t_0) > 0$ and $\phi(t_0) = \phi'(t_0) = 0$, we can apply Lemma 10 under the two constraints $z = o\left(\sqrt{\ell}\right)$ and $\ell < n/2$. The first one comes from the restriction on ε and will be fulfilled by our final choice for z. The second one comes from the restriction on t_0 and means that we want to discard at least half of the candidates what is actually the case for cryptanalytic applications. Thus, we will need to compute the three following integrals.

Lemma 11 *Let $a > 1$ be a real number, we have:*

1. $\int_0^a e^{-t} \cdot t^{-1/2}\, dt = \sqrt{\pi} - e^{-a}a^{-1/2} + O(e^{-a}a^{-3/2})$.
2. $\int_0^a e^{-t}\, dt = 1 - e^{-a}$.
3. $\int_0^a e^{-t} \cdot t^{1/2}\, dt = \frac{\sqrt{\pi}}{2} - e^{-a}\sqrt{a} + O(e^{-a}a^{-1/2})$.

Proof This is easily done using integration by parts. □

Hence, applying this to Lemma 10, we have:

$$\begin{aligned} \int_{t_0}^{t_0+\varepsilon} e^{-\lambda\phi(t)}\, dt = & \sqrt{\frac{\pi}{2\lambda\phi''(t_0)}} + O\left(\frac{e^{-\lambda\phi(t_0+\varepsilon)}}{\lambda\sqrt{\phi''(t_0)\phi(t_0+\varepsilon)}}\right) \\ & - \frac{1}{3\lambda}\frac{\phi^{(3)}(t_0)}{\phi''^2(t_0)} + O\left(\frac{1}{\lambda}\frac{\phi^{(3)}(t_0)}{\phi''^2(t_0)}e^{-\lambda\phi(t_0+\varepsilon)}\right) \\ & + \frac{A_{t_0}}{2\lambda}\sqrt{\frac{\pi}{\lambda}} + O\left(A_{t_0}\frac{e^{-\lambda\phi(t_0+\varepsilon)}}{\lambda}\sqrt{\phi(t_0+\varepsilon)}\right) \end{aligned}$$

and,

$$\begin{aligned}
\int_{t_0-\varepsilon}^{t_0} e^{-\lambda\phi(t)}\,dt &= \sqrt{\frac{\pi}{2\lambda\phi''(t_0)}} + O\left(\frac{e^{-\lambda\phi(t_0-\varepsilon)}}{\lambda\sqrt{\phi''(t_0)\phi(t_0-\varepsilon)}}\right) \\
&\quad + \frac{1}{3\lambda}\frac{\phi^{(3)}(t_0)}{\phi''^2(t_0)} + O\left(\frac{1}{\lambda}\frac{\phi^{(3)}(t_0)}{\phi''^2(t_0)}e^{-\lambda\phi(t_0-\varepsilon)}\right) \\
&\quad + \frac{A_{t_0}}{2\lambda}\sqrt{\frac{\pi}{\lambda}} + O\left(A_{t_0}\frac{e^{-\lambda\phi(t_0-\varepsilon)}}{\lambda}\sqrt{\phi(t_0-\varepsilon)}\right).
\end{aligned}$$

Summing the two integrals gives:

$$\begin{aligned}
\int_{t_0-\varepsilon}^{t_0+\varepsilon} e^{-\lambda\phi(t)}\,dt &= \sqrt{\frac{2\pi}{\lambda\phi''(t_0)}} + O\left(\frac{e^{-\lambda(\phi(t_0-\varepsilon)} + e^{-\lambda\phi(t_0+\varepsilon))}}{\lambda\varepsilon\phi''(t_0)}\right) \\
&\quad + O\left(\frac{1}{3\lambda}\frac{\phi^{(3)}(t_0)}{\phi''^2(t_0)}\left(e^{-\lambda\phi(t_0-\varepsilon)} + e^{-\lambda\phi(t_0+\varepsilon)}\right)\right) \\
&\quad + \frac{A_{t_0}}{\lambda}\sqrt{\frac{\pi}{\lambda}} + O\left(\frac{A_{t_0}\varepsilon}{\lambda}\sqrt{\phi''(t_0)}\left(e^{-\lambda\phi(t_0-\varepsilon)} + e^{-\lambda\phi(t_0+\varepsilon)}\right)\right) \\
&= \sqrt{\frac{2\pi}{\lambda\phi''(t_0)}}\cdot\left[1 + \sqrt{\phi''(t_0)}\frac{A_{t_0}}{\lambda}\right] \\
&\quad + O\left(\frac{1}{\lambda}\left[e^{-\lambda\phi(t_0-\varepsilon)} + e^{-\lambda\phi(t_0+\varepsilon)}\right]\left[\frac{1}{\varepsilon\phi''(t_0)} + \frac{\phi^{(3)}(t_0)}{3\phi''^2(t_0)} + A_{t_0}\varepsilon\sqrt{\phi''(t_0)}\right]\right).
\end{aligned}$$

We now substitute the real values for λ and the derivatives of ϕ.

$$\begin{aligned}
\int_{t_0-\varepsilon}^{t_0+\varepsilon} \tilde{g}(t)\,dt &= \int_{t_0-\varepsilon}^{t_0+\varepsilon} C_{n,\ell}e^{-(n-2)D(t_0||t)}\,dt \\
&= C_{n,\ell}\cdot\sqrt{\frac{2\pi t_0(1-t_0)}{n-2}}\cdot\left[1 + \frac{13t_0^2 - 13t_0 + 1}{12(n-2)t_0(1-t_0)}\right] + R \\
&= \frac{n-1}{n-2}\cdot\left[1 + \frac{13t_0^2 - 13t_0 + 1}{12(n-2)t_0(1-t_0)}\right] + R \\
&= \left[1 + \frac{1}{n-2}\right]\cdot\left[1 + \frac{13t_0^2 - 13t_0 + 1}{12(n-2)t_0(1-t_0)}\right] + R \\
&= 1 + \frac{13t_0^2 - 13t_0 + 1}{12(\ell-1)t_0} + O\left(\frac{1}{n}\right) + R.
\end{aligned}$$

With R equals to:

$$R = O\left(\frac{C_{n,\ell}}{n}\left(e^{-\lambda\phi(t_0-\varepsilon)} + e^{-\lambda\phi(t_0+\varepsilon)}\right)\left[\frac{t_0(1-t_0)}{\varepsilon} + \frac{2}{3}(2t_0-1) + \frac{13t_0^2 - 13t_0 + 1}{12t_0(1-t_0)}\cdot\varepsilon\right]\right)$$

The sum into the brackets is dominated by the first term which is of order $\sqrt{\ell}/z$ thus,

$$\begin{aligned} R &= O\left(\frac{\sqrt{\ell}C_{n,\ell}}{z\cdot n}\left[e^{-(n-2)D(t_0||t_0-\varepsilon)}+e^{-(n-2)D(t_0||t_0+\varepsilon)}\right]\right) \\ &= O\left(\frac{\sqrt{\ell}}{z\cdot n}C_{n,\ell}e^{-(n-2)D(t_0||t_0-\varepsilon)}\left[1+e^{-(n-2)\Delta_\varepsilon}\right]\right). \end{aligned}$$

Using Lemma 8 gives

$$R = O\left(\frac{e^{-z^2/2}}{z}\left[1+e^{-(n-2)\Delta_\varepsilon}\right]\right).$$

Then, applying Lemma 9 leads to

$$e^{-(n-2)\Delta_\varepsilon} = O\left(e^{-\frac{2z^3}{3\sqrt{\ell}}}\right).$$

Thus, $R = O\left(\frac{e^{-z^2/2}}{z}\right)$. To conclude this proof:

$$\begin{aligned} \int_{t_0-\varepsilon}^{t_0+\varepsilon} g(t)\,dt &= \left[1+\frac{13t_0^2-13t_0+1}{12(\ell-1)t_0}+O\left(\frac{1}{n}+\frac{e^{-z^2/2}}{z}\right)\right] \\ &\quad\cdot\left[1-\frac{1}{12(\ell-1)}+O\left(\frac{1}{n}+\frac{1}{\ell^2}\right)\right] \\ &= 1-(1-t_0)\frac{13t_0-1}{12t_0}\cdot\frac{1}{\ell-1}+O\left(\frac{1}{\ell^2}+\frac{1}{n}+\frac{e^{-z^2/2}}{z}\right) \\ &= 1+O\left(\frac{1}{\ell^2}+\frac{1}{n}+\frac{e^{-z^2/2}}{z}\right) \end{aligned}$$

□

7.6 Proof of Lemma 6

We recall that we want to show that

$$\sum_{i=F^{-1}(t_0-\varepsilon)}^{F^{-1}(t_0)-1} f_0(i) = O\left(\frac{zC_\lambda\delta}{\sqrt{\ell-1}}\right),$$

where $\delta = \sum_{i=0}^{F^{-1}(t_0)-1} f_0(i)$ and z is defined from ε by $\varepsilon = z\frac{\sqrt{\ell-1}}{n}$.

First, and to simplify formulae, we denote by B and B_ε the values $B \overset{\text{def}}{=} F^{-1}(t_0)$ and $B_\varepsilon \overset{\text{def}}{=} F^{-1}(t_0-\varepsilon)$. In the case $B = B_\varepsilon$, there is no term in the sum and thus, the lemma is proved. We now assume that $B \geq B_\varepsilon + 1$.

The proof of Lemma 6 is based on Lemma 3. Thus, we will use coefficients

$$\gamma_0 \overset{\text{def}}{=} \frac{(1-p_0)\cdot B}{p_0\cdot(N-B+1)} \text{ and } \gamma \overset{\text{def}}{=} \frac{(1-p)\cdot B}{p\cdot(N-B+1)}. \tag{25}$$

Some technical lemmas are required to prove Lemma 6. The first one is given by

Lemma 12 *If* $\frac{\ell-1}{n-2} \leq \frac{1}{4}$ *and* ε *is chosen of the form* $\varepsilon = z\frac{\sqrt{\ell-1}}{n}$ *with* $z = o(\sqrt{\ell})$ *as* ℓ *goes to infinity, then we have*

$$(B - B_\varepsilon)(\gamma - 1) = O\left(\frac{z}{\sqrt{\ell-1}}\right)$$

when ℓ *goes to infinity.*

Proof On the one hand, by using Lemma 3 and by splitting the sum as explained above Theorem 1, it can be easily derived that

$$\sum_{i=B+1}^{N} f(i) = \theta\left(\frac{f(B)}{1-1/\gamma}\right) = \theta\left(\gamma\frac{f(B)}{\gamma-1}\right) \tag{26}$$

On the other hand, still using Lemma 3,

$$(\gamma_-^{B-B_\varepsilon} - 1)\frac{f(B)}{\gamma_- - 1} \leq \sum_{i=B_\varepsilon+1}^{B} f(i) \tag{27}$$

with

$$\gamma_- \overset{\text{def}}{=} \frac{1-p}{p}\min\left(\frac{B}{N-B+1}, \frac{B_\varepsilon+2}{N-B_\varepsilon-1}\right)$$

From the fact that $\frac{\ell-1}{n-2}$ is smaller than $\frac{1}{4}$ we infer that $B > Np$ and for ℓ large enough $B_\epsilon > Np$ (this is the only reason why we choose $\frac{\ell-1}{n-2} \leq \frac{1}{4}$). Therefore for ℓ large enough we have $\gamma_- = \gamma$. From the hypothesis made on ε we know that $\sum_{i=B_\varepsilon+1}^{B} f(i) = o\left(\sum_{i=B+1}^{N} f(i)\right)$ as ℓ goes to infinity. This is only possible if $\gamma_-^{B-B_\varepsilon} - 1$ goes to zero as ℓ goes to infinity. This in turn implies that $\gamma_-^{B-B_\varepsilon} - 1 \sim (B - B_\varepsilon)(\gamma_- - 1)$ as ℓ goes to infinity. This also implies the same statement with γ replacing γ_- (since γ_- coincides with γ for ℓ large enough). Putting all these remarks together with (27) and (26) we obtain

$$\begin{aligned}
(B - B_\varepsilon)(\gamma - 1) &\underset{\ell\to\infty}{\sim} \gamma_-^{B-B_\varepsilon} - 1 \\
&= O\left(\sum_{i=B_\varepsilon+1}^{B} f(i)\frac{\gamma-1}{f(B)}\right) \\
&= O\left(\frac{\sum_{i=B_\varepsilon+1}^{B} f(i)}{\sum_{i=B+1}^{N} f(i)}\right)
\end{aligned}$$

Then, we express the sums appearing in this fraction as functions of ε and t_0

$$\sum_{i=B_\varepsilon+1}^{B} f(i) = F(F^{-1}(t_0)) - F(F^{-1}(t_0-\varepsilon)) = \varepsilon\left[1 + O\left(\frac{f(B_\varepsilon)}{\varepsilon}\right)\right],$$

and

$$\sum_{i=B+1}^{N} f(i) = 1 - F(F^{-1}(t_0)) = (1-t_0)\left[1 + O\left(\frac{f(B)}{1-t_0}\right)\right].$$

And we finally obtain

$$(B - B_\varepsilon)(\gamma - 1) = O\left(\frac{\varepsilon}{1-t_0}\left[1 + O\left(\frac{f(B_\varepsilon)}{\varepsilon}\right)\right]\right).$$

It is straightforward to check that $O\left(\frac{f(B_\varepsilon)}{\varepsilon}\right) = O\left(1\right)$. Substituting the values taken for ε and t_0 we obtain

$$(B - B_\varepsilon)(\gamma - 1) = O\left(\frac{z}{\sqrt{\ell - 1}}\right).$$

□

Proof We now prove Lemma 6.

We can apply Lemma 3 and derive from it the following expressions

$$\sum_{i=0}^{B-1} f_0(i) = \theta\left(\frac{f_0(i)}{1-\gamma_0}\right) \tag{28}$$

$$\sum_{i=B_\varepsilon}^{B} f_0(i) = \theta\left(\frac{(1-\gamma_0^{B-B_\varepsilon})f_0(i)}{1-\gamma_0}\right) \tag{29}$$

We observe now that

$$\begin{aligned} 1 - \gamma_0^{B-B_\epsilon} &= O\left((B - B_\epsilon)(1-\gamma_0)\right) \\ &= O\left(C_\lambda(\gamma - 1)(B - B_\varepsilon)\right) \\ &= O\left(C_\lambda \frac{z}{\sqrt{\ell - 1}}\right). \end{aligned}$$

By putting this back into Equation (29) and using that $\delta \overset{\text{def}}{=} \sum_{i=0}^{B-1} f_0(i)$ we finally obtain Lemma 3.

□