On the Data Complexity of Statistical Attacks Against Block Ciphers

Céline Blondeau and Benoît Gérard

INRIA project-team SECRET, France
{celine.blondeau, benoit.gerard}@inria.fr

Abstract. Many attacks on iterated block ciphers rely on statistical considerations using plaintext/ciphertext pairs to distinguish some part of the cipher from a random permutation. We provide here a simple formula for estimating the amount of plaintext/ciphertext pairs which is needed for such distinguishers and which applies to a lot of different scenarios (linear cryptanalysis, differential-linear cryptanalysis, differential/truncated differential/impossible differential cryptanalysis). The asymptotic data complexities of all these attacks are then derived. Moreover, we give an efficient algorithm for computing the data complexity accurately.
Keywords : statistical cryptanalysis, iterated block cipher, data complexity.

1 Introduction

Distinguishing attacks against block ciphers aim at determining whether a permutation corresponds to a permutation chosen uniformly at random from the set of all permutations or one of the permutations specified by a secret key. Any such attack against an iterated block cipher is a serious threat since it can usually be transformed into a key-recovery attack, e.g. by combining it with an exhaustive search for the last round key. We focus here on the case where the attacker has a certain amount of plaintext/ciphertext pairs from which he deduces N binary samples whose sum follows a binomial distribution of parameters (N, p) in the case of a random permutation and (N, p_*) in the other case. Such attacks are referred as *non-adaptative iterated attacks* by Vaudenay [Vau03]. The problem addressed by all these attacks is to determine whether a sample results from a binomial distribution of parameter p_* or p.

The variety of statistical attacks covers a huge number of possibilities for (p_*, p). For instance, in linear cryptanalysis [TCG92,Mat93,Mat94], p_* is close to $p = \frac{1}{2}$ while in differential cryptanalysis [BS91], p is small and p_* is quite larger than p. Explicit formulae for the data complexity are well-known in both cases but there is a lack of such formulae for hybrid

cases, for instance for truncated differential attacks where both p and p_* are small and p/p_* is close to one.

Selçuk sums up the problem in [Sel08]: to express error probabilities, one has to calculate tails of binomial distributions which are not easy to manipulate. It is desirable to use an approximation of them. Actually, in differential cryptanalysis [LMM91], the well-known formula for the data complexity is obtained by using a Poisson approximation for binomial law, leading to a number of chosen plaintexts n of the form:

$$n \approx \frac{1}{p_*}.$$

But this approximation holds for small p_* only. In linear cryptanalysis [Mat93], a Gaussian approximation provides

$$n \approx \frac{1}{(p_* - p)^2}.$$

1.1 Related work

Ideally, we would like to have an approximation that can be used on the whole space of parameters. Actually, error probabilities vary with the number of samples N as a product of a polynomial factor $Q(N)$ and an exponential factor $2^{-\Gamma N}$:

$$Q(N)2^{-\Gamma N}.$$

The asymptotic behavior of the exponent has been exhibited by Baignères, Junod and Vaudenay [Jun03,BJV04,BV08] by applying some classical results from statistics. However, for many statistical cryptanalyses, the polynomial factor is non negligible. To our best knowledge, 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. For instance, the deep analysis of the complexity of linear attacks due to Junod [Jun01,Jun03,JV03] 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_* and p are slightly

larger than in a differential cryptanalysis but the ratio p_*/p is closer to 1.

Hereafter we present both attacks on generalized Feistel network [Nyb96] defined in Appendix A.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 defined in the field $GF(2^8)$ by the power permutation $x \mapsto x^7$.

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 the probability

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

Let T be a partition of $GF(2^8)$ into cosets of the subfield $GF(2^4)$. If α 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 $T_1 = \alpha^7 + GF(2^4)$ and $T_2 = GF(2^4)$. Let

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

For ten rounds of this generalized Feistel network with good subkeys, the probability of the truncated differential characterized by (A, B) is

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

For a random permutation the probability function of the output is independent from the input. 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:

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

Notice that the probabilities given have been theoretically computed and that they take into account all the differential pathes.

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

In this paper we propose a general framework to compare the data complexity of different statistical attacks.

Section 2 recalls the statistical framework of distinguishing attacks. Section 3 compares the formula for binomial tails computation we use (involving Kullback-Leibler divergence) to those classically used. Then, Section 4 gives a general method to estimate the minimal pair *threshold/amount of data* that fits with the attack requirements (i.e., that achieves given error probabilities). Section 5 elaborates on results given in Section 3 to provide a good estimate of the required amount of data for some given error probabilities. This approximation is actually quite close to the exact value and an upper bound on the relative error is given. We deduce that comparing different statistical cryptanalyses reduces to computing the corresponding Kullback-Leibler divergences.

Finally, in Section 6, we expand Kullback-Leibler divergence with a Taylor series for some specific statistical cryptanalyses. We recover some well-known behaviors and find some new ones.

2 Hypothesis testing

Many (non-adaptive) statistical attacks based on distinguishers can be modeled in the following way. The attacker performs a guess on a subkey K of the cipher and wishes to know whether this guess is correct or not. There are two possibilities:

- H_{good}: "K is the correct guess".
- H_{bad}: "K is not the correct guess".

The attacker has a certain way of distinguishing the right subkey and a certain amount of plaintext/ciphertext pairs from which he is able to calculate N binary values $X_1, X_2, \ldots, X_N$ which are independent and identically distributed and satisfy

$$P(X_i = 1|H_{\text{good}}) = p_*,$$
$$P(X_i = 1|H_{\text{bad}}) = p.$$

From the samples $X_1, \ldots, X_N$ the attacker either decides that H_{good} holds or that H_{bad} is true. Two kind of errors are possible:

- **Non-detection:** It occurs if it is decided that there is a wrong subkey guess when H_{good} holds. We denote by α the non-detection error probability.

– **False alarm:** It occurs if one decides that K is the right subkey when H_{bad} holds. We denote by β the false alarm error probability.

By using well known results about hypothesis testing it follows that $\left\{\mathbf{X} \in \{0;1\}^N\ ,\ S_N = \sum_{i=1}^N X_i \geq T\right\}$ is an optimal acceptance region for some integer $0 \leq T \leq N$. The meaning of optimal is stated in the following lemma.

Lemma 1. *[CT91]**Neyman-Pearson lemma** :*
If distinguishing between two hypotheses H_{good} and H_{bad} with N samples $(X_1, \ldots, X_N)$ using a test of the form :

$$\frac{P(X_1, \ldots, X_N | H_{\text{good}})}{P(X_1, \ldots, X_N | H_{\text{bad}})} \geq t$$

gives error probabilities P_{nd} and P_{fa}, then no other test can improve both non-detection and false alarm error probabilities.

A standard calculus (detailed in [CT91] for the Gaussian case) shows that comparing the ratio of Lemma 1 with a real number t is equivalent to compare $S_N = \sum_{i=1}^N X_i$ with an integer $0 \leq T \leq N$.

3 Approximating error probabilities

This section introduces and compares different ways of approximating error probabilities. For the attacks we consider in this paper, computing those error probabilities amounts to computing binomial tails. A particular quantity will play a fundamental role here, the Kullback-Leibler divergence.

Definition 2. Kullback-Leibler divergence [CT91]
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:

$$D\left(p||q\right) = p \log_2 \left(\frac{p}{q}\right) + (1-p) \log_2 \left(\frac{1-p}{1-q}\right).$$

We use the convention (based on continuity arguments) that $0 \log_2 \frac{0}{p} = 0$ and $p \log_2 \frac{p}{0} = \infty$.

Later, we will denote by log the base 2 logarithm.

Our main tool is a theorem borrowed from [AG89] which captures exactly the exponential behavior of the binomial tails together with the right polynomial factor. Recall that $S_{N,p} = \sum_{i=1}^{N} X_i$ where the X_i's follow a Bernoulli distribution of parameter p.

Writing $f \underset{N\to\infty}{\sim} g$ means $\lim_{N\to\infty} \frac{f(N)}{g(N)} = 1$. The main result in [AG89] is the following theorem:

Theorem 1. *Let p_* and p be two real numbers such that $0 < p < p_* < 1$ and $0 < \tau < 1$. Then,*

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

and

$$P(S_{N,p_*} \leq \tau N) \underset{N\to\infty}{\sim} \frac{p_*\sqrt{1-\tau}}{(p_*-\tau)\sqrt{2\pi N\tau}} 2^{-ND(\tau||p_*)}. \tag{2}$$

We are now going to compare these estimates with the ones classically used.

In [BJV04,BV08], the aim of the authors is to derive an asymptotic formula for the best distinguisher, that is the distinguisher that maximizes $|1-\alpha-\beta|$. We denote by N the number of requests of the distinguisher. The following result is used:

$$\max(\alpha, \beta) \doteq 2^{-NC(p_*,p)} \tag{3}$$

where $f(N) \doteq g(N)$ means $f(N) = g(N)e^{o(N)}$.

In the general case where $p_* \notin \{0, 1\}$, such a distinguisher has an acceptance region of the form mentionned by Lemma 1 with t equals to 1. In this setting, the value of the relative threshold τ fulfills the equality $D(\tau||p_*) = D(\tau||p)$. Actually, this value of the Kullback-Leibler divergence is the Chernoff information $C(p_*, p)$ used by Junod, Baignères and Vaudenay (see [CT91, Section 12.9]). The exponent in (1) and (2) is the same that the one given by (3):

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

In the case $p_* = 0$ or $p_* = 1$, in impossible or higher order differential cryptanalysis for instance, the relative threshold τ is equal to p_* and the non-detection error probability α vanishes. Thus, $\max(\alpha, \beta) = \beta \doteq 2^{-ND(p_*||p)} \doteq 2^{-NC(p_*,p)}$. The last equality is directly derived from the

definition of the Kullback-Leibler divergence. So we also find the same exponent as in [BV08] in this particular case.

In [BJV04], a polynomial factor is taken into account but it is only suitable where the Gaussian approximation of binomial tails can be used. For instance, this formula gives a bad estimate in the case of differential cryptanalysis:

$$N \approx \frac{2 \cdot \Phi^{-1}(\frac{\alpha+\beta}{2})^2}{D\left(p_* || p\right)}, \tag{4}$$

where Φ^{-1} is the inverse cumulative function of a Gaussian random variable.

Hereafter we compare N (the required number of samples) to the estimates obtained using (3) and (4). The value of $\log(N)$ is obtained thanks to Algorithm 1 presented in Section 4 with some refinement detailed in Appendix A.5. The results are summed-up in Figure 1. An additional column contains the estimate found using (1) and (2). Note that the corresponding estimate tends towards N as β goes to zero.

To sum-up this section, asymptotic studies on distinguishers as [BV08] neglect the polynomial factor when approximating error probabilities. Obviously, such estimations overestimate the real complexity as shown in Figure 1.

In [BJV04,BV08] the authors take a threshold τ that maximize the advantage $|1 - \alpha - \beta|$. The maximum is obtained for two error probabilities α and β which are roughly the same. However, the time complexity of a cryptanalysis depends on β. Therefore, it is often the case that this probability is chosen to be much smaller than the non-detection probability.

We also observe that the approximation given in [BJV04] and Selçuk's one [Sel08] are tight when the Gaussian approximation is suitable but are rather poor everywhere else. In this paper, we fill this gap giving a unique formula using a polynomial factor that can be used for all sets of parameters p_* and p.

4 General method

In this section we use the previously defined notation. We are interested in finding an accurate number of samples to reach given error probabilities.

Let $S_{N,p}$ (resp. S_{N,p_*}) be a random variable which follows a binomial law of parameters N and p (resp. p_*). The acceptance region is defined by the threshold T, thus both error probabilities can be rewritten as $P_{nd} = P(S_{N,p_*} < T)$ and $P_{fa} = P(S_{N,p} \geq T)$. Let α and β be two given real numbers $(0 < \alpha, \beta < 1)$. The problem is to find a number of samples

		$\log(N)$	(1) & (2)	[BJV04]	[BV08]
Linear	$p_* = 0.5 + 1.49 \cdot 2^{-24}$ $p = 0.5$ $\alpha = 0.1$ $\beta = 0.1$	47.57	47.88	47.57	49.58
Linear	$p_* = 0.5 + 1.49 \cdot 2^{-24}$ $p = 0.5$ $\alpha = 0.001$ $\beta = 0.001$	50.10	50.13	50.10	51.17
Differential	$p_* = 1.87 \cdot 2^{-56}$ $p = 2^{-64}$ $\alpha = 0.1$ $\beta = 0.1$	56.30	56.77	54.44	57.71
Differential	$p_* = 1.87 \cdot 2^{-56}$ $p = 2^{-64}$ $\alpha = 0.001$ $\beta = 0.001$	58.30	58.50	56.98	59.29
Truncated differential	$p_* = 1.18 \cdot 2^{-16}$ $p = 2^{-16}$ $\alpha = 0.001$ $\beta = 0.001$	26.32	26.35	26.28	27.39

Fig. 1. Estimations of $\log(N)$ from [BJV04,BV08] and our work for some parameters.

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:

$$\begin{cases} P(S_{N,p_*} < T) \leq \alpha, \\ P(S_{N,p} \geq T) \leq \beta. \end{cases}$$

In practice, using real numbers avoids some 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 but one can use more accurate estimates using formulae given in Appendix A.5.

We respectively denote by $G_{\mathrm{nd}}(N, \tau)$ and $G_{\mathrm{fa}}(N, \tau)$ the estimates for non-detection and false alarm error probabilities.

In consequence, we want to find N and τ such that

$$G_{\mathrm{nd}}(N, \tau) \leq \alpha \quad \text{and} \quad G_{\mathrm{fa}}(N, \tau) \leq \beta. \tag{5}$$

For a given τ, G_{nd} and G_{fa} are essentially decreasing functions of N. This means that for a given τ, we can compute $N_{\mathrm{nd}}(\tau)$ and $N_{\mathrm{fa}}(\tau)$ the values 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 those 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 2. *Let $G_{\mathrm{nd}}(N,\tau)$ and $G_{\mathrm{fa}}(N,\tau)$ be two functions of N and τ, defined on $[0;+\infty] \times [p;p_*]$, 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;*
- $\lim_{N\to 0} G_{\mathrm{nd}}(N,\tau) \geq 1 \quad , \quad \lim_{N\to 0} G_{\mathrm{fa}}(N,\tau) \geq 1, \lim_{N\to\infty} G_{\mathrm{nd}}(N,\tau) = \lim_{N\to\infty} G_{\mathrm{fa}}(N,\tau) = 0$

Let us recall that for fixed α, β in $[0;1]$ and τ in $[p;p_*]$, $G_{\mathrm{nd}}(N_{\mathrm{nd}}(\tau),\tau)=\alpha$ 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 (5).

Then, for $p \leq m \leq p_*$,
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 $N(\tau) > N(m)$.

$\Diamond$

N_{nd} and N_{fa} can be found thanks to a dichotomic search but a more efficient way of doing that is explained in Appendix A.4.

Application. Our first motivation was to compare differential and truncated differential cryptanalyses of a generalized Feistel network. For the cipher described in Section 1 the results obtained with some fixed error probabilities are given in Figure 2. We recall that in the case of differential cryptanalysis $p = 2^{-32}$ and $p_* = 1.53 \times 2^{-27}$ while for truncated differential cryptanalysis, $p = 2^{-16}$ and $p_* = 1.18 \times 2^{-16}$.

This truncated differential cryptanalysis is thus an improvement of this differential one.

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

Input: Given error probabilities (α, β) and probabilities (p_*, p).
Output: N and τ: the minimum number of samples and the corresponding relative threshold to reach error probabilities less than (α, β).

Set τ_{min} to p and τ_{max} to p_*.
repeat
  Set τ to $\frac{\tau_{min} + \tau_{max}}{2}$.
  Compute N_{nd} such that $\forall N > N_{\mathrm{nd}}, G_{\mathrm{nd}}(N, \tau) \leq \alpha$.
  Compute N_{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 τ.

α	β	$\log(N)$ (differential)	$\log(N)$ (truncated differential)
0.5	0.001	27.35	24.31
0.5	10^{-10}	29.25	26.37
0.01	0.001	29.43	25.94
0.01	10^{-10}	30.54	27.29

Fig. 2. Number of required samples N for differential and truncated-differential cryptanalyses.

5 Asymptotic behavior

The aim of this section is to provide a simple criterion to compare two different statistical attacks. Such attacks rely on the fact that some phenomena are more likely to appear in the output of some secret key dependent permutation than in a random permutation. So an attack is defined by a pair (p_*, p) of probabilities where p (resp. p_*) is the probability of the phenomenon to occur in the random permutation output (resp. in a key dependent permutation output).

In order to simplify following calculus, we take a threshold $\tau = p_*$ that gives a non-detection error probability P_{nd} 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 $\tau = p_*$ 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_* (resp. p) be the probability of the phenomenon to occur in the key dependent permutation output (resp. the random permutation output). For a relative threshold $\tau = p_*$, a good approximation of the required number of samples N to distinguish between the key dependent permutation and the random permutation with false alarm error probability less or equal to β is*

$$N' = -\frac{1}{D\left(p_* || p\right)} \left[\log\left(\frac{\lambda\beta}{\sqrt{D\left(p_* || p\right)}} \right) + 0.5 \log\left(-\log(\lambda\beta) \right) \right], \quad (6)$$

since

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

for

$$\lambda = \frac{(p_* - p)\sqrt{2\pi(1 - p_*)}}{(1 - p)\sqrt{p_*}} \quad \text{and} \quad \theta = \left[1 + \frac{1}{2\log(\lambda\beta)} \log\left(-\frac{\log(\lambda\beta)}{D\left(p_* || p\right)} \right) \right]^{-1}. \quad (7)$$

Where N_∞ is the value obtained using Algorithm 1, (1) and (2).
Proof. See Appendix A.2. ◊

This approximation with N' is tight : we estimated the data complexity of some known attacks (see Figure 3) and observed θ's in the range $]1; 6.5]$. Moreover, for $\beta = 2^{-32}$, observed values of θ's were less than 2.

A simple comparison for statistical attacks

Equation (6) gives a simple way of roughly comparing the data complexity of two statistical attacks. Indeed, N' is essentially a decreasing function of $D(p_*||p)$. Therefore, comparing the data complexity of two statistical cryptanalyses boils down to comparing the Kullback-Leibler divergences of those cryptanalyses. .

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

$$N'' = -\frac{\log(2\sqrt{\pi}\beta)}{D(p_*||p)}. \tag{8}$$

Experimental results given in Section 7 show that this estimation is quite sharp and becomes better when 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 estimations 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.

6 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 $D(p_*||p)^{-1}$. Hereafter, we estimates $D(p_*||p)^{-1}$ for many statistical cryptanalyses. We recover the format of known results and give new results for truncated differential and higher order differential cryptanalysis. Let us recall the Kullback-Leibler divergence

$$D(p_*||p) = p_* \log\left(\frac{p_*}{p}\right) + (1-p_*)\log\left(\frac{1-p_*}{1-p}\right).$$

In Appendix A.3, Lemma 3 gives an estimate of Kullback-Leibler divergence

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

Linear cryptanalysis. In the case of linear cryptanalysis, p_* is close to $p = 1/2$. Thus we get

$$\frac{1}{D\left(p_*||p\right)} \approx \frac{1}{(p_* - p)^2}.$$

If we use the notation of linear cryptanalysis ($p_* - p = \varepsilon$), we recover ε^{-2}, which is a well-known result due to Matsui [Mat93,Mat94].

Differential cryptanalysis. In this case, both p_* and p are small but the difference $p_* - p$ is dominated by p_*.

$$\frac{1}{D\left(p_*||p\right)} \approx \frac{1}{p_* \log(p_*/p) - p_*}.$$

This result is slightly different from the commonly used result, e.g. $\frac{1}{p_*}$ in [LMM91] because it involves $\log(p_*/p)$. However, the commonly used result requires some restrictions on the ratio p_*/p so it is natural that such a dependency appears.

Differential-linear cryptanalysis. This attack presented in [LH94] combines a 3-round differential characteristic of probability 1 with a 3-round linear approximation. This gives $p = 0.5$ and $p_* = 0.576$. This case is very similar to linear cryptanalysis since we observe a linear behavior in the output. Thus, as it is written in [LH94], the asymptotic behavior of the number of samples is

$$\frac{1}{D\left(p_*||p\right)} \approx \frac{1}{(p_* - p)^2}.$$

Truncated differential cryptanalysis. In the case of truncated differential cryptanalysis, p_* and p are small but close to each other. This leads to

$$\frac{1}{D\left(p_*||p\right)} \approx \frac{p}{(p_* - p)^2}.$$

Impossible differential. This case is a particular one. The impossible differential cryptanalysis [BBS99] relies on the fact that some event cannot occur in the output of the key dependent permutation. We have always assumed that $p_* > p$ but in this case it is not true anymore ($p_* = 0$). However, the formula holds in this case too:

$$\frac{1}{D\left(0||p\right)} = \log^{-1}\left(\frac{1}{1-p}\right) \approx p^{-1}.$$

Higher order differential. This attack introduced in [Knu94] is a generalization of differential cryptanalysis. It exploits the fact that a k-th order differential of the cipher is constant (i.e independent from the plaintext and the key). A typical case is when $k = deg(F+1))$, any k-th order differential of F vanishes. Therefore, for this attack, we have $p_* = 1$. Moreover, $p = (2^m - 1)^{-1}$ where m is the block size so p is small.

$$\frac{1}{D\left(1||p\right)} = \log^{-1}\left(\frac{1}{p}\right) = -1/\log(p).$$

An important remark here, is that in a cryptanalysis of order k, a sample corresponds to 2^k chosen plaintexts.

7 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 Section 5, namely:

$$N' = -\frac{1}{D\left(p_*||p\right)}\left[\log\left(\frac{\lambda\beta}{\sqrt{D\left(p_*||p\right)}}\right) + 0.5\log\left(-\log(\lambda\beta)\right)\right] \text{ with } \lambda = \frac{(p_* - p)\sqrt{2\pi(1-p_*)}}{(1-p)\sqrt{p_*}},$$

and,

$$N'' = \frac{-\log(2\sqrt{\pi}\beta)}{D\left(p_*||p\right)}.$$

In Figure 3, 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.

In statistical cryptanalysis, we extract the key of the cipher in a list of candidates for the good key. The smaller the false alarm probability is, the smaller the list of candidates will be. And we can see in Figure 3 that when β goes to 0, N' and N'' tend to N.

		p	p_*	$\log(N)$	$\log(N')$	$\log(N'')$
$\beta = 2^{-8}$	**L**	0.5	$0.5 + 1.19 \cdot 2^{-21}$	42.32	42.00 (-0.32)	42.60 ($+0.28$)
	DL	0.5	$0.5 + 1.73 \cdot 2^{-6}$	11.26	11.15 (-0.11)	11.52 ($+0.26$)
	D	2^{-64}	$1.87 \cdot 2^{-56}$	54.57	54.68 ($+0.11$)	54.82 ($+0.25$)
	Dgfn	2^{-32}	$1.53 \cdot 2^{-27}$	27.14	26.80 (-0.34)	26.94 (-0.20)
	TDgfn	2^{-16}	$1.18 \cdot 2^{-16}$	23.85	23.66 (-0.19)	24.13 ($+0.28$)
$\beta = 2^{-16}$	**L**	0.5	$0.5 + 1.19 \cdot 2^{-21}$	43.62	43.54 (-0.08)	43.79 ($+0.17$)
	DL	0.5	$0.5 + 1.73 \cdot 2^{-6}$	12.54	12.52 (-0.02)	12.71 ($+0.17$)
	D	2^{-64}	$1.87 \cdot 2^{-56}$	55.85	55.94 ($+0.09$)	56.02 ($+0.17$)
	Dgfn	2^{-32}	$1.53 \cdot 2^{-27}$	28.27	28.05 (-0.22)	28.14 (-0.13)
	TDgfn	2^{-16}	$1.18 \cdot 2^{-16}$	25.15	25.11 (-0.04)	25.33 ($+0.18$)
$\beta = 2^{-32}$	**L**	0.5	$0.5 + 1.19 \cdot 2^{-21}$	44.78	44.76 (-0.02)	44.88 ($+0.10$)
	DL	0.5	$0.5 + 1.73 \cdot 2^{-6}$	13.70	13.69 (-0.01)	13.80 ($+0.10$)
	D	2^{-64}	$1.87 \cdot 2^{-56}$	56.98	57.06 ($+0.08$)	57.11 ($+0.13$)
	Dgfn	2^{-32}	$1.53 \cdot 2^{-27}$	29.13	29.17 ($+0.04$)	29.23 ($+0.10$)
	TDgfn	2^{-16}	$1.18 \cdot 2^{-16}$	26.31	26.30 (-0.01)	26.42 ($+0.11$)

Fig. 3. Some experiments for some values of parameters β, p and p_*.

The parameters p_* and p considered are :

- **L** : DES linear cryptanalysis recovering 26 key bits [Mat94].
- **DL** : DES differential-linear cryptanalysis [LH94].
- **D** : DES differential cryptanalysis [BS93].
- **Dgfn** : Generalized Feistel networks differential cryptanalysis presented in this paper.
- **TDgfn** : Generalized Feistel networks truncated differential cryptanalysis presented in this paper.

8 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 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 this formula (2) is a decreasing function of $D(p_*||p)^{-1}$. This implies that comparing the data complexity of different statistical cryptanalyses boils down to computing the corresponding Kullback-Leibler divergences. Actually, the behavior of the number of samples is dominated by $D(p_*||p)^{-1}$. We show that $D(p_*||p)^{-1}$ gives the same order of magnitude as known results excepted in differential cryptanalysis where a dependency on $\log(p_*/p)$ is emphasized. We also extend these results to other block cipher statistical cryptanalyses, for instance, truncated differential cryptanalysis. To conclude, Figure 4 sums up the behaviors of the number of required samples for some known statistical cryptanalyses. Some experimental results are given in Section 7 to compare estimates found in Section 5 to the real value of N. These results show the accuracy of the estimates given in Section 5 in the settings of actual cryptanalyses.

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

Fig. 4. Asymptotic data complexity for some statistical attacks.

References

[AG89] R. Arriata and L. Gordon. Tutorial on large deviations for the binomial distribution. *Bulletin of Mathematical Biology*, 51(1):125–131, 1989.

[BBS99] E. Biham, A. Biryukov, and A. Shamir. Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differentials. In *EUROCRYPT '99*, volume 1592 of *LNCS*, pages 12–23, 1999.

[BDK02] E. Biham, O. Dunkelman, and N. Keller. Enhancing Differential-Linear Cryptanalysis. In *ASIACRYPT '02*, volume 2501 of *LNCS*, pages 254–266. SV, 2002.

[BJV04] T. Baignères, P. Junod, and S. Vaudenay. How Far Can We Go Beyond Linear Cryptanalysis? In *ASIACRYPT '04*, volume 3329 of *LNCS*, pages 432–450. Springer–Verlag, 2004.

[BS91] E. Biham and A. Shamir. Differential Cryptanalysis of DES-like Cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991.

[BS93] E. Biham and A. Shamir. Differential Cryptanalysis of the Full 16-round DES. In *CRYPTO'92*, volume 740 of *LNCS*, pages 487–496. Springer–Verlag, 1993.

[BV08] T. Baignères and S. Vaudenay. The Complexity of Distinguishing Distributions. In *ICITS '08*, volume 5155 of *LNCS*, pages 210–222. SV, 2008.

[CT91] T.M. Cover and J.A. Thomas. *Information theory*. Wiley series in communications. Wiley, 1991.

[Gil97] H. Gilbert. *Cryptanalyse statistique des algorithmes de chiffrement et sécurité des schémas d'authentification*. PhD thesis, Université Paris 11 Orsay, 1997.

[Jun01] P. Junod. On the Complexity of Matsui's Attack. In *SAC '01*, volume 2259 of *LNCS*, pages 199–211. Springer–Verlag, 2001.

[Jun03] P. Junod. On the Optimality of Linear, Differential, and Sequential Distinguishers. In *EUROCRYPT '03*, volume 2656 of *LNCS*, pages 17–32. Springer–Verlag, 2003.

[Jun05] P. Junod. *Statistical cryptanalysis of block ciphers*. PhD thesis, EPFL, 2005.

[JV03] P. Junod and S. Vaudenay. Optimal key ranking procedures in a statistical cryptanalysis. In *FSE '03*, volume 2887 of *LNCS*, pages 235–246. Springer–Verlag, 2003.

[Knu94] L. R. Knudsen. Truncated and Higher Order Differentials. In *FSE '94*, volume 1008 of *LNCS*, pages 196–211. Springer–Verlag, 1994.

[LH94] S. K. Langford and M. E. Hellman. Differential-Linear Cryptanalysis. In *CRYPTO '94*, volume 839 of *LNCS*, pages 17–25. Springer–Verlag, 1994.

[LMM91] X. Lai, J. L. Massey, and S. Murphy. Markov Ciphers and Differential Cryptanalysis. *LNCS*, 547:17–38, 1991.

[Mat93] M. Matsui. Linear cryptanalysis method for DES cipher. In *EUROCRYPT '93*, volume 765 of *LNCS*, pages 386–397. Springer–Verlag, 1993.

[Mat94] M. Matsui. The First Experimental Cryptanalysis of the Data Encryption Standard. In *CRYPTO '94*, volume 839 of *LNCS*, pages 1–11. Springer–Verlag, 1994.

[Nyb96] K. Nyberg. Generalized Feistel Networks. In *ASIACRYPT '96*, volume 1163 of *LNCS*, pages 91–104. Springer–Verlag, 1996.

[Sel08] A. A. Selçuk. On Probability of Success in Linear and Differential Cryptanalysis. *Journal of Cryptology*, 21(1):131–147, 2008.

[TCG92] A. Tardy-Corfdir and H. Gilbert. A Known Plaintext Attack of FEAL-4 and FEAL-6. In *CRYPTO '91*, volume 576 of *LNCS*, pages 172–181. Springer–Verlag, 1992.

[Vau03] S. Vaudenay. Decorrelation: A Theory for Block Cipher Security. *Journal of Cryptology*, 16(4):249–286, 2003.

A Appendix

A.1 Generalized Feistel networks

A generalized Feistel network [Nyb96] is an iterated block cipher whose round function is depicted in Figure 5.

Definition 3. *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:*

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

where $\oplus$ is the modulo 2 addition.

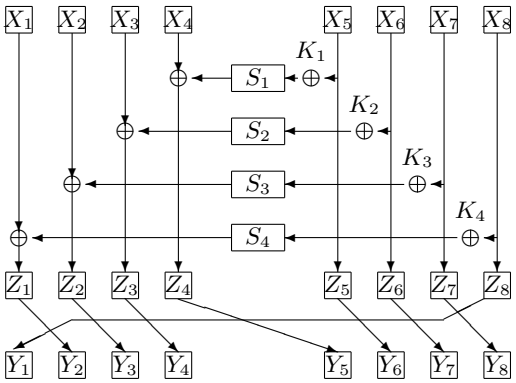

Fig. 5. Generalized Feistel network with 4 S-boxes

A.2 Proof of Theorem 2

Proof. Recall that $\tau = p_*$ 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 (1) in Theorem 1 gives

$$N \approx -\frac{\log(\lambda\beta\sqrt{N})}{D\left(p_* || p\right)} \tag{9}$$

where

$$\lambda = \frac{(p_* - p)\sqrt{2\pi(1-p_*)}}{(1-p)\sqrt{p_*}}.$$

Formula (9) suggests to bring in the following function f which has a contraction property

$$f(x) = -\frac{\log(\lambda\beta\sqrt{x})}{D\left(p_*||p\right)}.$$

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:

$$f(x) = a - b\log(x),$$

with

$$a = -\frac{\log(\lambda\beta)}{D\left(p_*||p\right)} \text{ and } b = \frac{1}{2D\left(p_*||p\right)}.$$

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.

$$\begin{aligned} N_3 = a - b\log(N_2) &= N_1 - b\log\left(N_1 \cdot \frac{N_2}{N_1}\right) \\ &= N_1 - b\log(N_1) + b\log\left(\frac{N_1}{N_2}\right) \\ &= N_2 + b\log\left(\frac{N_1}{N_2}\right) \end{aligned}$$

Let us define

$$\theta = \left[1 + \frac{1}{2\log(\lambda\beta)}\log\left(-\frac{\log(\lambda\beta)}{D\left(p_*||p\right)}\right)\right]^{-1},$$

as in Equation (7) in Theorem 2. Then,

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

The bound on N_∞ becomes:

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

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

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

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

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

$\Diamond$

A.3 Taylor expansion of the Kullback-Leibler divergence

Lemma 3. *Let $0 < p < p_* < 1$. Then,*

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

Proof.

Using the Taylor theorem, we get

$$(1 - p_*) \log\left(\frac{1 - p_*}{1 - p}\right) = p - p_* + \frac{(p - p_*)^2}{2(1 - p_*)} + O(p - p_*)^3.$$

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

$\Diamond$

A.4 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 P_{nd} to α, (2) can be rewritten as:

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

Using the same fixed point argument as in Appendix A.2, we can find N_{nd} by iterating the function with a first point $x_0 = D\left(\tau||p_*\right)^{-1}$. The same thing can be done with (1) in order to find N_{fa}.

A.5 Discussion on Algorithm 1: Accurate computation of error probabilities

To accurately estimate error probabilities, we use Stirling approximation of the binomial coefficient :

$$\binom{a}{b} \simeq \frac{1}{\sqrt{2\pi}}\sqrt{\frac{a}{(a-b)b}}\frac{a^a}{(a-b)^{a-b}b^b} \tag{10}$$

If S_{N,p_*} follows a binomial distribution of parameters N and p_*,

$$P(S_{N,p_*} = T-1) = \frac{1-p_*}{p_*} \cdot \frac{T}{N-T+1} \cdot P(S_{N,p_*} = T).$$

This leads to :

$$\begin{aligned}
P(S_{N,p_*} < T) &= P(S_{N,p_*} = T) \cdot \left[\frac{(1-p_*)\cdot T}{p_* \cdot (N-T+1)} + \frac{(1-p_*)^2 \cdot T(T-1)}{p_*^2 \cdot (N-T+1)(N-T+2)} + \cdots\right] \\
&= P(S_{N,p_*} = T) \cdot \sum_{i=1}^{T} \left(\frac{1-p_*}{p_*}\right)^i \frac{T!}{(T-i)!}\frac{(N-T)!}{(N-T+i)!} \\
&= \binom{N}{T} p_*^T (1-p_*)^{N-T} \cdot \sum_{i=1}^{T} \left(\frac{1-p_*}{p_*}\right)^i \frac{T!}{(T-i)!}\frac{(N-T)!}{(N-T+i)!}.
\end{aligned}$$

From (10), we estimate the probability :

$$\begin{aligned}
P(S_{N,p_*} = T) &= \binom{N}{T} p_*^T (1-p_*)^{N-T} \\
&\simeq \frac{1}{\sqrt{2\pi}} \sqrt{\frac{N}{(N-T)T}} \frac{N^N}{(N-T)^{N-T} T^T} p_*^T (1-p_*)^{N-T} \\
&\simeq \sqrt{\frac{N}{2\pi(N-T)T}} \left(\frac{Np_*}{T}\right)^T \left(\frac{(1-p_*)N}{N-T}\right)^{N-T} \\
&\simeq \sqrt{\frac{N}{2\pi(N-T)T}} 2^{-N\left(\frac{T}{N}\log(\frac{T}{N}/p_*)+(1-\frac{T}{N})\log((1-\frac{T}{N})/(1-p_*))\right)} \\
&\simeq \sqrt{\frac{N}{2\pi(N-T)T}} 2^{-ND(\frac{T}{N}||p_*)}
\end{aligned}$$

Finally, we get :

$$P(S_{N,p_*} < T) = P(S_{N,p_*} \leq T-1) \simeq \frac{2^{-ND(\frac{T}{N}||p_*)}}{\sqrt{2\pi(1-\frac{T}{N})T}} \cdot \sum_{i=1}^{T} \left(\frac{1-p_*}{p_*}\right)^i \frac{T!}{(T-i)!} \frac{(N-T)!}{(N-T+i)!}. \tag{11}$$

The key is to notice that the dominant term is the last one. So, we begin to sum with this term and then add the others until it reaches a given precision. This estimate is tight when N and T are big enough. When T is small, one can use the exact formula of binomial probability since the complexity comes from the size of T. The same thing can be done for false alarm error probability.

We would like to use this estimation in Algorithm 1 to compute a very good estimate of N. Algorithm 1 uses a formula of error probabilies with relative threshold τ because of its continuity. We have to extend (11) to real numbers. For a relative threshold τ, we want a formula for $P(S_{N,p_*} < \tau N)$ that corresponds to (11) when τN is an integer and that is continous. Let T_{up} be the value $\lceil \tau N \rceil$. Then such an estimate is the following :

$$P(S_{N,p_*} < \tau N) \simeq \frac{2^{-ND\left(\frac{T_{up}}{N}||p_*\right)}}{\sqrt{2\pi(1-\frac{T_{up}}{N})T_{up}}} \cdot \left[1 - (T_{up} - \tau N) + \sum_{i=1}^{T_{up}} \left(\frac{1-p_*}{p_*}\right)^i \frac{T_{up}!}{(T_{up}-i)!} \frac{(N-T_{up})!}{(N-T_{up}+i)!}\right]. \tag{12}$$

We can derive from (12) an expression with contraction properties to compute N_{nd} in Algorithm 1. Error made on the estimation can be

bounded because the error on Stirling approximation is well-known and the error when not summing until T_{up} can be roughly bounded using:

$$P(S_{N,p_*} = i - j) \leq \left(\frac{1 - p_*}{p_*} \cdot \frac{i}{N - i + 1} \right)^j \cdot P(S_{N,p_*} = i).$$

All this work can be done for false alarm probability.