The Complexity of Distinguishing Distributions

Thomas Baignères* and Serge Vaudenay

EPFL
CH-1015 Lausanne, Switzerland
http://lasecwww.epfl.ch

Abstract. Cryptography often meets the problem of distinguishing distributions. In this paper we review techniques from hypothesis testing to express the advantage of the best distinguisher limited to a given number of samples. We link it with the Chernoff information and provide a useful approximation based on the squared Euclidean distance. We use it to extend linear cryptanalysis to groups with order larger than 2.[1]

1 Preliminaries

1.1 Best Distinguisher

The hypothesis testing problem can be considered as a simple game in which a first player uses a *source* to generate independent random samples in some given finite set $\mathcal{Z}$ with a distribution P which follows either a *null hypothesis* H_0 or an alternate hypothesis H_1. The second player, often called *distinguisher*, must determine which hypothesis was used by using the samples. In the simplest testing problem, the source follows a distribution $\mathsf{P} \in \{\mathsf{P}_0, \mathsf{P}_1\}$ chosen among two distributions, both being known to the distinguisher. He faces two hypotheses, namely $\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$ and $\mathsf{H}_1 : \mathsf{P} = \mathsf{P}_1$. This situation is commonly referred to as the *simple hypothesis testing problem* since both alternatives fully determine the distribution. A more complex situation arises when one of the two hypotheses is *composite*, i.e., when the distinguisher has to guess whether the distribution followed by the source is one particular distribution ($\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$) or if it belongs to a set of several distributions ($\mathsf{H}_1 : \mathsf{P} \in \{\mathsf{P}_1, \ldots, \mathsf{P}_d\}$). Finally, the difficulty of the game can be increased from the point of view of the distinguisher if the exact description of the alternate hypothesis is not available. In that case, it shall guess whether the source follows a specific (known) distribution ($\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$) or not ($\mathsf{H}_1 : \mathsf{P} \neq \mathsf{P}_0$).

* Supported by the Swiss National Science Foundation, 200021-107982/1

[1] These results will be part of [1].

In all cases, the adversary is assumed to be computationally unbounded[2] and to be only limited by the number q of samples available, so that we will referred to it as a q-limited distinguisher and denote it A_q. If $\boldsymbol{Z}^q = Z_1, \ldots, Z_q$ are the q samples available to A_q, we define the type I error α and the type II error β by:

$$\alpha = \Pr_{\mathsf{H}_0}[\mathsf{A}_q(\boldsymbol{Z}^q) = 1] \qquad \beta = 1 - \Pr_{\mathsf{H}_1}[\mathsf{A}_q(\boldsymbol{Z}^q) = 1]$$

For composite hypotheses, these probabilities make sense when distributions are assigned weights (following the Bayesian approach). We measure the ability to distinguish between hypothesis H_0 and H_1 by the *advantage* defined as

$$\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{H}_0, \mathsf{H}_1) = \left|\Pr_{\mathsf{H}_0}[\mathsf{A}_q(\boldsymbol{Z}^q) = 1] - \Pr_{\mathsf{H}_1}[\mathsf{A}_q(\boldsymbol{Z}^q) = 1]\right| = |1 - \alpha - \beta|\,.$$

In the simple hypothesis case we denote the advantage by $\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{P}_0, \mathsf{P}_1)$. We let

$$\mathsf{P}_{\boldsymbol{Z}^q}[z] = \frac{n_z}{q},$$

be the relative proportion of occurrences of each symbol of $\mathcal{Z}$ ($\mathsf{P}_{\boldsymbol{Z}^q}$ is also called *type* of $\boldsymbol{Z}^q$ [4]), where n_z is the number of occurrences of the symbol z in the sequence $\boldsymbol{Z}^q = Z_1, \ldots, Z_q$. Since the samples are assumed to be mutually independent, their particular order must be irrelevant. Consequently, the final distinguishing decision can be solely based on the type $\mathsf{P}_{\boldsymbol{Z}^q}$ of the sequence. Denoting $\mathcal{P}$ the set of all probability distributions over $\mathcal{Z}$, we can completely describe any distinguisher by an *acceptance region* $\Pi \subset \mathcal{P}$ such that

$$\mathsf{A}_q(\boldsymbol{Z}^q) = 1 \quad \Leftrightarrow \quad \mathsf{P}_{\boldsymbol{Z}^q} \in \Pi.$$

For $q = 1$ we can easily show (see [2]) that $\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{P}_0, \mathsf{P}_1)$ reaches a maximum equal to

$$\mathrm{BestAdv}_1(\mathsf{P}_0, \mathsf{P}_1) = \frac{1}{2}\|\mathsf{P}_0 - \mathsf{P}_1\|_1$$

where the norm $\|\cdot\|_1$ of a function f is defined by $\|f\|_1 = \sum_x |f(x)|$. We can apply this result to the probability distribution of $\boldsymbol{Z}^q$. By using the equality

$$2(aa' - bb') = (a - b)(a' + b') + (a' - b')(a + b)$$

[2] So that we can assume w.l.o.g. that the adversary is fully deterministic.

we deduce that

$$\mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1) \leq \frac{q}{2}\|\mathsf{P}_0 - \mathsf{P}_1\|_1$$

The first concern of the present paper is to obtain a more precise expression for $\mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1)$.

Notations. The natural logarithm is denoted ln while log refers to basis 2 logarithm. The *support* of a distribution P is the set $\mathsf{Supp}(\mathsf{P})$ of all z for which $\mathsf{P}[z] > 0$. In this paper, P_0 and P_1 will be two distinct distributions on a finite set $\mathcal{Z}$ such that $\mathsf{Supp}(\mathsf{P}_0) \cup \mathsf{Supp}(\mathsf{P}_1) = \mathcal{Z}$. We will denote $\mathcal{Z}' = \mathsf{Supp}(\mathsf{P}_0) \cap \mathsf{Supp}(\mathsf{P}_1)$. In the case where both P_0 and P_1 are of full support we have $\mathcal{Z} = \mathcal{Z}'$, otherwise $\mathcal{Z}' \subsetneq \mathcal{Z}$. The *Chernoff information*[3] between P_0 and P_1 is

$$\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = -\inf_{0<\lambda<1} \log \sum_{z\in\mathcal{Z}'} \mathsf{P}_0[z]^{1-\lambda}\mathsf{P}_1[z]^{\lambda}.$$

The Kullback-Leibler divergence between P_0 and P_1 is

$$\mathrm{D}(\mathsf{P}_0\|\mathsf{P}_1) = \sum_{z\in\mathsf{Supp}(\mathsf{P}_0)} \mathsf{P}_0[z] \log \frac{\mathsf{P}_0[z]}{\mathsf{P}_1[z]}$$

with the convention that $\mathrm{D}(\mathsf{P}_0\|\mathsf{P}_1) = +\infty$ when $\mathsf{Supp}(\mathsf{P}_0) \not\subseteq \mathsf{Supp}(\mathsf{P}_1)$. The notation $f(q) \doteq g(q)$ for $q \to +\infty$ means that $f(q) = g(q)e^{o(q)}$ or equivalently that

$$\lim_{q\to+\infty} \frac{1}{q} \log \frac{f(q)}{g(q)} = 0.$$

We denote $f(q) \sim g(q)$ for $f(q) = g(q)(1 + o(1))$.

1.2 Neyman-Pearson

Given 3 distributions $\mathsf{P}_0, \mathsf{P}_1, \mathsf{P}$, let us define

$$\mathrm{L}(\mathsf{P}) = \sum_{z\in\mathsf{Supp}(\mathsf{P})} \mathsf{P}[z] \log \frac{\mathsf{P}_0[z]}{\mathsf{P}_1[z]}$$

with the natural convention that $\log 0 = -\infty$ and $\frac{1}{0} = +\infty$. (Note that if P has a support either included in the one of P_0 or in the one of

[3] Note that our definition differs from that sometimes given (e.g., in [4, p.314]), namely $\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = -\min_{0\leq\lambda\leq1} \log \sum_{z\in\mathcal{Z}} \mathsf{P}_0[z]^{1-\lambda}\mathsf{P}_1[z]^{\lambda}$, since the latter is not well defined when $\mathsf{Supp}(\mathsf{P}_0) \neq \mathsf{Supp}(\mathsf{P}_1)$.

P_1 then we never encounter an illegal operation such as $\frac{0}{0}$ or $\infty - \infty$.) The best distinguisher between P_0 and P_1 can be expressed as follows. Given a sample vector $\boldsymbol{Z}^q$ we compute $\mathrm{L}(\mathsf{P}_{\boldsymbol{Z}^q})$ (which is nothing but the logarithmic likelihood ratio). The distinguisher is defined by a threshold τ and outputs 1 iff $\mathrm{L}(\mathsf{P}_{\boldsymbol{Z}^q}) \leq \log \tau$. The Neyman-Pearson Lemma [6] says that for any distinguisher achieving error probabilities α and β, there exists τ such that the above distinguisher has error probabilities not larger than α and β respectively. This means that for any distinguisher there exists one based on the likelihood ratio which is at least as good in terms of error probabilities.

If one is concerned with maximizing the advantage (or equivalently in minimizing $\alpha + \beta$) then the best distinguisher is defined by $\tau = 1$. It can be defined by the acceptance region

$$\Pi = \{\mathsf{P} \in \mathcal{P} \ : \ \mathrm{L}(\mathsf{P}) \leq 0\}.$$

A classical result (see [4, Section 12.9]) gives a precise asymptotic expression for α and β when P_0 and P_1 have the *same support*.

Theorem 1. *Let P_0 and P_1 be two distributions of finite support $\mathcal{Z}$. Let $\mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1)$ denote the best advantage for distinguishing P_0 from P_1 with q samples and α and β the type I and type II errors of the distinguisher, respectively. We have*

$$1 - \mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1) \doteq \alpha \doteq \beta \doteq 2^{-q\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1)}.$$

Unfortunately, this expression of α and β is not correct if the supports do not match as the following example shows.

Example 2. We can consider $\mathcal{Z} = \{1, 2, 3\}$ and

$$\mathsf{P}_0 = \left(\tfrac{1}{3}\ \tfrac{1}{3}\ \tfrac{1}{3}\right) \qquad \mathsf{P}_1 = \left(a\ b\ 0\right)$$

with $a + b = 1$, $\frac{1}{3} > a > \frac{1}{7}$. We have

$$\mathrm{L}(\mathsf{P}) = \begin{cases} \mathsf{P}[1] \log \frac{1}{3a} + \mathsf{P}[2] \log \frac{1}{3b} & \text{if } \mathsf{P}[3] = 0 \\ +\infty & \text{if } \mathsf{P}[3] \neq 0 \end{cases}$$

The Chernoff information is computed from the minimum over $]0, 1[$ of

$$F(\lambda) = \frac{1}{3}(3a)^\lambda + \frac{1}{3}(3b)^\lambda.$$

This is a convex function such that $F(0) = \frac{2}{3}$ and $F(1) = 1$. Assuming that $a \in]\frac{1}{3}, \frac{1}{7}[$, since $a + b = 1$ we have $9ab > 1$ thus $F'(0) > 0$. We

deduce that F is increasing over $]0,1[$ so the minimum is $F(0) = \frac{2}{3}$: we have $\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = -\log \frac{2}{3}$. Since $F(\lambda) \to +\infty$ when $\lambda \to -\infty$ the minimum of F is reached for some $\lambda < 0$ which we call λ_0. We have $2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_1)} = \left(\frac{2}{3}\right)^q$. The type I error α is the probability that $\mathrm{L}(\mathsf{P}_{\boldsymbol{Z}^q}) \leq 0$ under distribution P_0 which mandates that 3 never occurs. This holds with probability $\left(\frac{2}{3}\right)^q$. When this happens, the number of occurrences of 1 and 2 are roughly similar so $\mathrm{L}(\mathsf{P}_{\boldsymbol{Z}^q}) \leq 0$. We can indeed show that $\alpha \doteq \left(\frac{2}{3}\right)^q$, which matches the result of Theorem 1. However, the type II error β is the probability that $\mathrm{L}(\mathsf{P}_{\boldsymbol{Z}^q}) > 0$ under distribution P_1 which is the probability that $n_1 \log 3a + n_2 \log 3b < 0$. This means that 2 must occur much less than 1 although its probability b is higher than a. As a consequence of Theorem 3 below we can show that $\beta \doteq F(\lambda_0)^q$ which does not match Theorem 1. The expression is thus correct for α but incorrect for β. In what follows we show that the expression is always correct for $\max(\alpha, \beta)$ so it is still correct for the advantage.

2 Best Advantage for Simple Hypothesis Testing

2.1 Result

Theorem 3. *Let P_0 and P_1 be two distributions of finite supports with union $\mathcal{Z}$ and intersection $\mathcal{Z}'$. Given a distribution P over $\mathcal{Z}$ we define*

$$\mathrm{L}(\mathsf{P}) = \sum_{z \in \mathsf{Supp}(\mathsf{P})} \mathsf{P}[z] \log \frac{\mathsf{P}_0[z]}{\mathsf{P}_1[z]} \quad \textit{and} \quad F(\lambda) = \sum_{z \in \mathcal{Z}'} \mathsf{P}_0[z]^{1-\lambda} \mathsf{P}_1[z]^{\lambda}.$$

Let $\Pi = \{\mathsf{P} \in \mathcal{P} \ : \ \mathrm{L}(\mathsf{P}) \leq 0\}$ be the acceptance region of the best distinguisher. Its type I error α satisfies

$$\alpha \doteq \left(\inf_{\lambda > 0} F(\lambda)\right)^q.$$

If there exists $z \in \mathcal{Z}'$ such that $0 < \mathsf{P}_1[z] < \mathsf{P}_0[z]$ then

$$\beta \doteq \left(\inf_{\lambda < 1} F(\lambda)\right)^q.$$

Otherwise, $\beta = 0$.

If for all $z \in \mathcal{Z}'$ we have $\mathsf{P}_1[z] \geq \mathsf{P}_0[z]$ then β is clearly zero and $\inf_{\lambda>0} F(\lambda) = F(0)$ so $\max(\alpha, \beta) = \alpha \doteq 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_1)}$. Otherwise, we note that

$$\max\left(\inf_{]0,+\infty[} F, \inf_{]-\infty,1[} F\right) = \inf_{]0,1[} F$$

because F is a convex function. Hence, we still have

$$\max(\alpha, \beta) \doteq 2^{-q\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1)}.$$

We deduce the following result.

Corollary 4. *Let P_0 and P_1 be two distributions of finite support with intersection $\mathcal{Z}'$. We have*

$$1 - \mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1) \doteq 2^{-q\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1)} = \left(\inf_{0<\lambda<1} \sum_{z \in \mathcal{Z}'} \mathsf{P}_0[z]^{1-\lambda} \mathsf{P}_1[z]^{\lambda} \right)^q.$$

2.2 Proof of Theorem 3

We first recall Sanov's theorem. To do this, we recall some notions of topology. The set of all functions from the finite set $\mathcal{Z}$ to $\mathbf{R}$ is a vector space of finite dimension thus all norms $\|\cdot\|$ define the same topology. An *open set* is a union of open balls, i.e. a set of functions f satisfying $\|f - f_0\| < r$ for a given function f_0 and a given radius $r \in \mathbf{R}$. The *interior* of a set Π is the union $\overset{\circ}{\Pi}$ of all open sets included in Π. A *closed set* is an intersection of closed balls. The *closure* of a set Π is the intersection $\overline{\Pi}$ of all closed sets containing Π.

Theorem 5 (Sanov [7]). *Let P_0 be a distribution over a finite set $\mathcal{Z}$ and $\boldsymbol{Z}^q = Z_1, \ldots, Z_q$ be q mutually independent random variables following distribution P_0. Let Π be a set of distributions over $\mathcal{Z}$ such that $\overline{\overset{\circ}{\Pi}} = \overline{\Pi}$. We have*

$$\Pr[\mathsf{P}_{\boldsymbol{Z}^q} \in \Pi] \doteq 2^{-q\mathrm{D}(\Pi \| \mathsf{P}_0)}$$

where $\mathrm{D}(\Pi \| \mathsf{P}_0) = \inf_{\mathsf{P} \in \Pi} \mathrm{D}(\mathsf{P} \| \mathsf{P}_0)$.

Intuitively, the $\overline{\overset{\circ}{\Pi}} = \overline{\Pi}$ assumption means that Π has no isolated point which could substantially influence $\mathrm{D}(\mathsf{P} \| \mathsf{P}_0)$ but would exceptionally (if ever) be reached by $\mathsf{P}_{\boldsymbol{Z}^q}$.

Lemma 6. *Let P_0 be a distribution of finite support $\mathcal{Z}$. Let g be a function such that $g(z) > 0$ for all $z \in \mathcal{Z}$. Given a distribution P over $\mathcal{Z}$ we define*

$$\mathrm{L}(\mathsf{P}) = \sum_{z \in \mathsf{Supp}(\mathsf{P})} \mathsf{P}[z] \log \frac{\mathsf{P}_0[z]}{g(z)} \quad \textit{and} \quad F(\lambda) = \sum_{z \in \mathcal{Z}} \mathsf{P}_0[z]^{1-\lambda} g(z)^{\lambda}.$$

Let Π *be the set of distributions over* $\mathcal{Z}$ *such that* $\mathrm{L}(\mathsf{P}) \leq 0$ *and consider the distinguisher* A_q *who accepts* $\boldsymbol{Z}^q$ *(i.e., returns 1) iff* $\mathsf{P}_{\boldsymbol{Z}^q} \in \Pi$*. We have*

$$\Pr[\mathsf{A}_q(\boldsymbol{Z}^q) = 1] = \Pr[\mathsf{P}_{\boldsymbol{Z}^q} \in \Pi] \doteq \left(\inf_{\lambda>0} F(\lambda)\right)^q.$$

If Π *is now the set of all distributions such that* $\mathrm{L}(\mathsf{P}) < 0$ *and there exists* z *such that* $0 < \mathsf{P}_0[z] < g(z)$ *the result still holds. Otherwise, the probability is zero.*

Proof. We first assume that $\mathsf{P}_0[z] \geq g(z)$ for all z. If Π is defined by $\mathrm{L}(\mathsf{P}) \leq 0$, the probability is $\mathsf{P}_0(\mathcal{Z}'')^q$ where $\mathcal{Z}''$ is the set of all z's such that $\mathsf{P}_0[z] = g(z)$, and the result easily comes. If Π is defined by $\mathrm{L}(\mathsf{P}) < 0$, the probability is clearly zero.

We now assume that we have $0 < \mathsf{P}_0[z] < g(z)$ for some z. Clearly, the distribution P such that $\mathsf{P}(z) = 1$ verifies $\mathrm{L}(\mathsf{P}) < 0$ so Π is nonempty. Considering the topology of distributions over $\mathcal{Z}$, we notice that L is continuous. Since $\mathrm{L}(\mathsf{P}) < 0$ for some $\mathsf{P} \in \Pi$, for $\varepsilon > 0$ small enough all distributions within a distance to P smaller than ε are in Π as well. This means that the interior of Π is nonempty. We note that Π is a convex set. Consequently, we have $\overline{\overset{\circ}{\Pi}} = \overline{\Pi}$ so that Sanov's theorem applies and we have

$$\Pr[\mathsf{P}_{\boldsymbol{Z}^q} \in \Pi] \doteq 2^{-q\mathrm{D}(\Pi\|\mathsf{P}_0)}.$$

What remains to be shown is that $\mathrm{D}(\Pi\|\mathsf{P}_0)$ is equal to $-\inf_{\lambda>0} \log F(\lambda)$ for both possible definitions of Π.

The set $\overline{\Pi}$ is bounded and topologically closed in a real vector space of finite dimension and therefore compact. We notice that $\mathsf{P} \mapsto \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$ is continuous on $\overline{\Pi}$. We deduce that $\mathrm{D}(\Pi\|\mathsf{P}_0) = \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$ for some P in $\overline{\Pi}$: we do have global minima for this function in $\overline{\Pi}$. Furthermore, the function $\mathsf{P} \mapsto \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$ is convex since

$$\mathrm{D}((1-t)\mathsf{P} + t\mathsf{P}'\|\mathsf{P}_0) \leq (1-t)\mathrm{D}(\mathsf{P}\|\mathsf{P}_0) + t\mathrm{D}(\mathsf{P}'\|\mathsf{P}_0)$$

so we deduce that there is no local minimum which is not global as well. Since the set of P's such that $\mathrm{D}(\mathsf{P}\|\mathsf{P}_0) \leq r$ is a convex set for any radius r, the set of global minima is indeed a convex set as well. Finally, if P reaches a minimum, then the segment between P_0 and P except P contains distributions "closer" (in the sense of D) to P_0 which must then be outside of Π. Thus their L value are positive. So, either the segment is reduced to P_0 (meaning that $\mathrm{L}(\mathsf{P}_0) \leq 0$) or we must have $\mathrm{L}(\mathsf{P}) = 0$ due to the continuity of L. Hence, the closest P in Π is either P_0 (if $\mathsf{P}_0 \in \overline{\Pi}$) or some P such that $\mathrm{L}(\mathsf{P}) = 0$.

We consider the differentiable function $\mathsf{P} \mapsto \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$ over the open space $\{\mathsf{P} : \mathcal{Z} \longrightarrow \mathbf{R}_+^*\}$ with constraints $\mathrm{N}(\mathsf{P}) = 1$ and $\mathrm{L}(\mathsf{P}) = \mathsf{cste}$ where $\mathrm{N}(\mathsf{P}) = \sum_z \mathsf{P}(z)$. By looking at the differentials, we have

$$\frac{\partial \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)}{\partial \mathsf{P}(a)} = \log \frac{\mathsf{P}(a)}{\mathsf{P}_0[a]} + \frac{1}{\ln 2}$$

so $\frac{\partial^2 \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)}{\partial \mathsf{P}(a)\partial \mathsf{P}(b)} = 0$ for $a \neq b$ and is strictly positive otherwise. Hence the second differential of $\mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$ is a strictly positive quadratic form. Thus, P is a local minimum for $\mathrm{D}(\cdot\|\mathsf{P}_0)$ over the distributions whose L value is constant iff the first differential is a linear combination of $d\mathrm{N}$ and $d\mathrm{L}$. This is the case iff P is of form P_λ for some λ where

$$\mathsf{P}_\lambda[z] = \frac{\mathsf{P}_0[z]^{1-\lambda} g(z)^\lambda}{\sum_a \mathsf{P}_0[a]^{1-\lambda} g(a)^\lambda}.$$

We deduce that for all $\lambda \in \mathbf{R}$, P_λ is *the* closest (in the sense of D) distribution to P_0 with this $\mathrm{L}(\mathsf{P}_\lambda)$ value. We look for the one for which this is zero.

We observe that F is a convex function such that $F(0) = 1$ and $F'(0) = -\mathrm{L}(\mathsf{P}_0)\ln 2$. More precisely, we have $F'(\lambda) = -\mathrm{L}(\mathsf{P}_\lambda)F(\lambda)\ln 2$. Since there exists z such that $\mathsf{P}_0[z] < g(z)$ the limit of F at $+\infty$ is $+\infty$. We note that

$$\mathrm{D}(\mathsf{P}_\lambda\|\mathsf{P}_0) = -\lambda \mathrm{L}(\mathsf{P}_\lambda) - \log F(\lambda).$$

If the closest P is not P_0 we have $\mathrm{L}(\mathsf{P}_0) \geq 0$ hence $F'(0) \leq 0$, so there must be a $\lambda \geq 0$ such that $F'(\lambda) = 0$ and for which $F(\lambda)$ is minimal. Clearly, this minimum is $\inf_{\lambda>0} F(\lambda)$. We deduce $\mathrm{L}(\mathsf{P}_\lambda) = 0$ thus P_λ is the closest distribution to P_0 in Π. The above expression of the distance yields the announced result in this case.

When P_0 is in $\overline{\Pi}$ we have $\mathrm{L}(\mathsf{P}_0) \leq 0$ thus $F'(0) \geq 0$. Since F is convex, F is increasing on $[0, +\infty[$ so $\inf_{\lambda>0} F(\lambda) = F(0) = 1$. Since $0 = D(\Pi\|\mathsf{P}_0)$ the result holds in this case as well. □

Proof (of Theorem 3). Let $\tilde{\mathsf{P}}_0[z] = \mathsf{P}_0[z]/\mathsf{P}_0(\mathcal{Z}')$ for $z \in \mathcal{Z}'$ and $\tilde{\mathsf{P}}_0[z] = 0$ otherwise. Let $g(z) = \mathsf{P}_1[z]/\mathsf{P}_0(\mathcal{Z}')$ for $z \in \mathcal{Z}'$ and $g(z) = 0$ otherwise. Applying Lemma 6 to $\tilde{\mathsf{P}}_0$ and g over $\mathcal{Z}'$ defines two functions $\tilde{\mathrm{L}}$ and $\tilde{F}$ and a set $\tilde{\Pi}$ of distributions over $\mathcal{Z}'$ satisfying $\tilde{\mathrm{L}}(\mathsf{P}) \leq 0$. Clearly, we have $\tilde{\mathrm{L}}(\mathsf{P}) = \mathrm{L}(\mathsf{P})$ for any distribution over $\mathcal{Z}'$. Indeed, Π consists of $\tilde{\Pi}$ plus all the distributions of support included in the one of P_1 but not in $\mathcal{Z}'$. The probability to reach one of these latter distributions when sampling z's following P_0 is clearly zero. Hence, the probability of accepting $\boldsymbol{Z}^q$ is the

probability that $\mathsf{P}_{\boldsymbol{Z}^q} \in \tilde{\Pi}$, under H_0. It is $\mathsf{P}_0(\mathcal{Z}')^q$ times the probability that $\mathsf{P}_{\boldsymbol{Z}^q} \in \tilde{\Pi}$ when sampling the Z_i's according to $\tilde{\mathsf{P}}_0$. By applying Theorem 1 we immediately obtain the result. □

3 Approximations for "Close" Distributions

We assume in this section that P_1 is *close* to P_0 of full support $\mathcal{Z}$. More precisely, we assume that P_0 is fixed of support $\mathcal{Z}$ and that P_1 tends towards P_0. Eventually, both distributions have the same support $\mathcal{Z}$, and for all $z \in \mathcal{Z}$ we have $x_z = o(1)$ as $\mathsf{P}_1 \to \mathsf{P}_0$ where

$$x_z = \frac{\mathsf{P}_1[z] - \mathsf{P}_0[z]}{\mathsf{P}_0[z]}.$$

3.1 Computing the Chernoff Information

Theorem 7. *Let P_0 be a distribution of support $\mathcal{Z}$. If the distribution P_1 over $\mathcal{Z}$ tends towards P_0, then*

$$\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) \sim \frac{1}{8 \ln 2} \sum_{z \in \mathcal{Z}} \frac{(\mathsf{P}_1[z] - \mathsf{P}_0[z])^2}{\mathsf{P}_0[z]}.$$

Proof. We let $x = (x_z)_{z \in \mathcal{Z}}$ and consider

$$F(\lambda, x) = \sum_{z \in \mathcal{Z}} \mathsf{P}_0[z](1 + x_z)^\lambda$$
$$g(\lambda, x) = \sum_{z \in \mathcal{Z}} \mathsf{P}_0[z](1 + x_z)^\lambda \ln(1 + x_z).$$

We define $\lambda^* \in [0, 1]$ as the value verifying $g(\lambda^*, x) = 0$. In terms of λ, $F(\lambda, x)$ is strictly convex of derivative $g(\lambda, x)$. Clearly, $\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = -\log F(\lambda^*, x)$. We will approximate $F(\lambda^*, x)$ when x is small and subject to $\sum_z \mathsf{P}_0[z] x_z = 0$. We first have

$$g(\lambda, x) = \sum_z \mathsf{P}_0[z](1 + \lambda x_z + o(x_z)) \left(x_z - \frac{x_z^2}{2} + o(x_z^2) \right)$$
$$= \sum_z \mathsf{P}_0[z] \left(\lambda - \frac{1}{2} \right) x_z^2 + o\left(\|x\|_2^2 \right)$$

since $\sum_z \mathsf{P}_0[z] x_z$ is zero. As $g(\lambda^*, x) = 0$ we deduce that λ^* tends towards $\frac{1}{2}$ as x tends towards 0. We now let

$$F(\lambda^*, x) = F\left(\frac{1}{2}, x \right) + \left(\lambda^* - \frac{1}{2} \right) F'_\lambda \left(\frac{1}{2}, x \right) + \frac{1}{2} \left(\lambda^* - \frac{1}{2} \right)^2 R$$

with $|R| \leq \max_\lambda F''_\lambda(\lambda, x)$ for $\lambda \in [0,1]$. As $F'_\lambda(\lambda, x) = g(\lambda, x)$, previous computations immediately lead to $F'_\lambda(\frac{1}{2}, x) = g(\frac{1}{2}, x) = o(\|x\|_2^2)$. Similarly we have

$$\begin{aligned} F''_\lambda(\lambda, x) &= \sum_{z \in \mathcal{Z}} \mathsf{P}_0[z](1+x_z)^\lambda \left(\ln(1+x_z)\right)^2 \\ &= \sum_{z \in \mathcal{Z}} \mathsf{P}_0[z](1+o(1))\left(x_z + o(x_z)\right)^2 \\ &= \sum_{z \in \mathcal{Z}} \mathsf{P}_0[z]x_z^2 + o(\|x\|^2) \end{aligned}$$

which is a $O(\|x\|^2)$, hence

$$F(\lambda^*, x) = F\left(\frac{1}{2}, x\right) + o(\|x\|^2).$$

Now, we have

$$\begin{aligned} F\left(\frac{1}{2}, x\right) &= \sum_{z \in \mathcal{Z}} \mathsf{P}_0[z]\sqrt{1+x_z} \\ &= \sum_{z \in \mathcal{Z}} \mathsf{P}_0[z]\left(1 + \frac{1}{2}x_z - \frac{1}{8}x_z^2 + o(x_z^2)\right) \\ &= 1 - \frac{1}{8}\sum_{z \in \mathcal{Z}} \mathsf{P}_0[z]x_z^2 + o(\|x\|_2^2) \end{aligned}$$

and therefore

$$F(\lambda^*, x) = 1 - \frac{1}{8}\sum_{z \in \mathcal{Z}} \mathsf{P}_0[z]x_z^2 + o(\|x\|_2^2),$$

which can be written

$$F(\lambda^*, x) = 1 - \frac{1}{8}\sum_{z \in \mathcal{Z}} \frac{(\mathsf{P}_1[z] - \mathsf{P}_0[z])^2}{\mathsf{P}_0[z]} + o\left(\sum_{z \in \mathcal{Z}}\left(\frac{\mathsf{P}_1[z] - \mathsf{P}_0[z]}{\mathsf{P}_0[z]}\right)^2\right).$$

□

Our computations were based on the assumption that $x_z = o(1)$ for all z. In practice however, both distribution are fixed. Yet we can use Theorem 7 to approximate $\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1)$ when $|\mathsf{P}_1[z] - \mathsf{P}_0[z]| \ll \mathsf{P}_0[z]$ for all z.

3.2 Close-to-Uniform Distributions

In the particular case where P_0 is the uniform distribution over $\mathcal{Z}$ of cardinality n, Theorem 7 yields

$$\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) \sim \frac{n}{8 \ln 2} \|\mathsf{P}_1 - \mathsf{P}_0\|_2^2$$

for the Euclidean norm $\|\cdot\|_2$, which can be used as the approximation

$$\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) \approx \frac{n}{8 \ln 2} \|\mathsf{P}_1 - \mathsf{P}_0\|_2^2$$

when $|\mathsf{P}_1[z] - \frac{1}{n}| \ll \frac{1}{n}$ for all z. When $\mathcal{Z}$ has a group structure, this can be expressed as

$$\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) \approx \frac{1}{8 \ln 2} \|\hat{\mathsf{P}}_1 - \hat{\mathsf{P}}_0\|_2^2 \quad \text{or even} \quad \mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) \approx \frac{1}{8 \ln 2} \sum_{\chi \in \hat{Z}} \mathsf{LP}(\chi)$$

where $\hat{\mathcal{Z}}$ is the dual group of $\mathcal{Z}$ (i.e., the set of all group homomorphisms χ between $\mathcal{Z}$ and the non-zero complex numbers) and where $\hat{\mathsf{P}}$ is the Fourier transform of P, i.e.

$$\hat{\mathsf{P}}(\chi) = \sum_{z \in \mathcal{Z}} \mathsf{P}[z] \bar{\chi}(z) \quad \text{and} \quad \mathsf{LP}(\chi) = |\hat{\mathsf{P}}_1(\chi)|^2 = |\mathrm{E}(\chi(Z))|^2$$

where Z follows the distribution P_1. This formally proves a heuristic result from Baignères, Stern, and Vaudenay [3] by showing that the best advantage is approximately

$$1 - e^{-\frac{q}{8} \|\hat{\mathsf{P}}_1 - \hat{\mathsf{P}}_0\|_2^2}$$

for q large and $\|\hat{\mathsf{P}}_1 - \hat{\mathsf{P}}_0\|_2$ small.

4 A Case of Composite Hypothesis Testing

So far, we considered the problem of testing the null hypothesis $\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$ against the simple alternate hypothesis $\mathsf{H}_1 : \mathsf{P} = \mathsf{P}_1$ where P_0 and P_1 were fully specified. We now consider the problem of distinguishing the case where P is equal to a specified distribution P_0 (the null hypothesis H_0) from the case where P belongs to a set $\mathcal{D} = \{\mathsf{P}_1, \ldots, \mathsf{P}_d\}$ of d fully specified distributions (the hypothesis H_1). Under H_1 we assume that the selection of P_i is taken with an *a priori* weight of π_i to define the advantage for distinguishing H_0 from H_1. For simplicity we assume that all distributions have the same support $\mathcal{Z}$.

4.1 Complex Hypothesis Testing

Theorem 8. *Let P_0 be a distribution of support $\mathcal{Z}$ and $\mathcal{D} = \{\mathsf{P}_1, \ldots, \mathsf{P}_d\}$ be a finite set of distributions of support $\mathcal{Z}$. In order to test the null hypothesis $\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$ against $\mathsf{H}_1 : \mathsf{P} \in \mathcal{D}$, the advantage of the best q-limited distinguisher is such that*

$$1 - \mathrm{BestAdv}_q(\mathsf{P}_0, \mathcal{D}) \doteq \max_{1 \le i \le d} 2^{-q\mathrm{C}(\mathsf{P}_0, \mathsf{P}_i)}.$$

It is reached by the distinguisher accepting $\boldsymbol{Z}^q$ iff

$$\min_{1 \le i \le d} \sum_{z \in \mathcal{Z}} \mathsf{P}_{\boldsymbol{Z}^q}[z] \log \frac{\mathsf{P}_0[z]}{\mathsf{P}_i[z]} \le 0.$$

Proof. Consider a q-limited distinguisher A_q defined by an acceptance region Π and denote by Adv_q its advantage. We have

$$1 - \mathrm{Adv}_q = \Pr_{\mathsf{H}_0}[\mathsf{A}_q(\boldsymbol{Z}^q) = 1] + \sum_{i=1}^{d} \pi_i \Pr[\mathsf{A}_q(\boldsymbol{Z}^q) = 0 | \mathsf{P} = \mathsf{P}_i]$$

thus $1 - \mathrm{Adv}_q$ is at least the average of all $1 - \mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_i)$ with weight π_i, which are (asymptotically) $2^{-q\mathrm{C}(\mathsf{P}_0, \mathsf{P}_i)}$. We deduce that

$$1 - \mathrm{Adv}_q \dot{>} \sum_{i=1}^{d} \pi_i 2^{-q\mathrm{C}(\mathsf{P}_0, \mathsf{P}_i)} \quad \text{thus} \quad 1 - \mathrm{Adv}_q \dot{>} \max_{1 \le i \le d} 2^{-q\mathrm{C}(\mathsf{P}_0, \mathsf{P}_i)}.$$

We define

$$\mathrm{L}_i(\mathsf{P}) = \sum_{z \in Z} \mathsf{P}[z] \log \frac{\mathsf{P}_0[z]}{\mathsf{P}_i[z]}$$

and consider the distinguisher based on the likelihood ratio between P_0 and P_i which is the closest to $\mathsf{P}_{\boldsymbol{Z}^q}$. We have

$$\mathrm{D}(\mathsf{P} \| \mathsf{P}_i) = \sum_{z \in \mathcal{Z}} \mathsf{P}[z] \log \frac{\mathsf{P}[z]}{\mathsf{P}_i[z]}$$

so that $\mathrm{D}(\mathsf{P} \| \mathsf{P}_i) \le \mathrm{D}(\mathsf{P} \| \mathsf{P}_j)$ is equivalent to $\mathrm{L}_i(\mathsf{P}) \le \mathrm{L}_j(\mathsf{P})$. Finally, this distinguisher is based on $\mathrm{L}(\mathsf{P}) = \min_i \mathrm{L}_i(\mathsf{P})$ and accepts H_1 iff $\mathrm{L}(\mathsf{P}) \le 0$. Let Π_i be the set of all P's such that $\mathrm{L}_i(\mathsf{P}) \le 0$ and Π be the union of all Π_i's. The best distinguishers simply checks whether $\mathsf{P}_{\boldsymbol{Z}^q} \in \Pi$.

Looking at the proof of Theorem 1, we can first see that the probability that $\mathsf{P}_{\boldsymbol{Z}^q} \in \Pi$ under the null hypothesis is equivalent to $2^{-q\mathrm{D}(\Pi \| \mathsf{P}_0)}$ which is the maximum of $2^{-q\mathrm{D}(\Pi_i \| \mathsf{P}_0)}$, itself equal to $2^{-q\mathrm{C}(\mathsf{P}_i, \mathsf{P}_0)}$. We deduce that

$$\Pr_{\mathsf{H}_0}[\mathsf{A}_q(\boldsymbol{Z}^q) = 1] \doteq \max_{1 \le i \le d} 2^{-q\mathrm{C}(\mathsf{P}_0, \mathsf{P}_i)}.$$

When the Z_i's are sampled according to P_i under hypothesis H_1, the probability of rejection is the probability that $\mathsf{P}_{\boldsymbol{Z}^q} \notin \Pi$. This is less than the probability that $\mathsf{P}_{\boldsymbol{Z}^q} \notin \Pi_i$ and we know that it is equivalent to $2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_i)}$. Since this is less than the maximum of $2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_j)}$, the advantage Adv_q is such that

$$1 - \mathrm{Adv}_q \doteq \max_{1\le i\le d} 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_i)}$$

Therefore, this distinguisher has the best advantage, asymptotically. □

4.2 Example: Generalized Linear Cryptanalysis

Let X be a random variable over G, an Abelian group. Let χ be a character over G such that the group $\mathcal{Z} = \chi(\mathsf{G})$ is of order d. Let $Z = \chi(X)$. Let P_0 be the uniform distribution over $\mathcal{Z}$. For each $u \in \mathcal{Z}$ we consider the distribution P_u defined by $\mathsf{P}_u[u] = \frac{1-\varepsilon}{d} + \varepsilon$ and $\mathsf{P}_u[z] = \frac{1-\varepsilon}{d}$ for all $z \in \mathcal{Z}$ such that $z \neq u$. Note that $\mathsf{LP}(\chi) = \varepsilon^2$ when Z follows distribution P_u for any u. These distributions have the property that $\hat{\mathsf{P}}_u$ is flat in the sense that for all $\varphi \neq 1$, $|\hat{\mathsf{P}}_u(\varphi)| = \varepsilon$. In linear cryptanalysis [3,5], χ is the product of several characters with "independent" biased distributions. It thus inherits of a distribution P such that $\hat{\mathsf{P}}$ is the product of "independent" Fourier transforms (this is the Piling-up Lemma) and is flattened as such. We have the following result.

Theorem 9. *If $Z = \chi(X)$ where χ is a character of order d, the best distinguisher between the null hypothesis that Z is uniformly distributed in the range of χ and the alternate hypothesis that Z follows some distribution P_u with u unknown is defined by*

$$\mathsf{A}_q(\boldsymbol{Z}^q) = 1 \Leftrightarrow \max_u \mathsf{P}_{\boldsymbol{Z}^q}[u] \ge \frac{\log(1-\varepsilon)}{\log(1-\varepsilon) - \log(1+(d-1)\varepsilon)}$$

where the right-hand side is approximated by $\frac{1}{d}(1 + (d-1)\frac{\varepsilon}{2})$. This distinguisher has an advantage such that $1 - \mathrm{Adv}_q \doteq 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_1)}$ which is approximated by

$$1 - \mathrm{Adv}_q \approx e^{-q\frac{d-1}{8}\varepsilon^2}.$$

Proof. We use the distinguisher which outputs 1 iff $\min_u \mathrm{L}_u(\mathsf{P}_{\boldsymbol{Z}^q}) \le 0$ (as suggested by Theorem 8). Clearly, $\min_u \mathrm{L}_u$ is reached for the value of u which maximizes $\mathsf{P}_{\boldsymbol{Z}^q}[u]$. We obtain that $\boldsymbol{Z}^q$ is accepted iff

$$\max_u \mathsf{P}_{\boldsymbol{Z}^q}[u] \ge \frac{\log(1-\varepsilon)}{\log(1-\varepsilon) - \log(1+(d-1)\varepsilon)}$$

which is approximately $\frac{1}{d}(1+(d-1)\frac{\varepsilon}{2})$. As it is surprising enough, we stress that the best distinguisher is based on $\|\mathsf{P}_{\boldsymbol{Z}^q}\|_\infty$ and not on the statistical average of $\chi(X)$ as one would expect. We can now focus on its advantage. By Theorem 8 we have

$$1-\mathrm{Adv}_q \doteq \max_{u\in\mathcal{Z}} 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_u)}.$$

Since all $\mathrm{C}(\mathsf{P}_0,\mathsf{P}_u)$ are equal, we can focus on $\mathrm{C}(\mathsf{P}_0,\mathsf{P}_1)$. Assuming that $\varepsilon \ll \frac{1}{d}$, we obtain

$$\mathrm{C}(\mathsf{P}_0,\mathsf{P}_1) = -\inf_\lambda \log \frac{1}{d}\left((1+(d-1)\varepsilon)^\lambda + (d-1)(1-\varepsilon)^\lambda\right) \approx \frac{d-1}{8\ln 2}\,\varepsilon^2.$$

The advantage is thus roughly $1-e^{-\frac{d-1}{8}q\varepsilon^2}$. □

Another problem consists in distinguishing the null hypothesis that Z is uniformly distributed in the range of χ from the alternate hypothesis that Z follows some arbitrary distribution of known *flatness* ζ. We define the flatness of a distribution P_1 by $\|\hat{\mathsf{P}}_1-\hat{\mathsf{P}}_0\|_2$. (Previously, we had $\zeta=\varepsilon\sqrt{d-1}$.) For such distributions, the Chernoff information is approximated by $\mathrm{C}(\mathsf{P}_0,\mathsf{P}_1)\approx\frac{\zeta^2}{8\ln 2}$. By Theorem 8, the best distinguisher satisfies $1-\mathrm{Adv}_q \approx e^{-\frac{q}{8}\zeta^2}$. It is defined by accepting $\boldsymbol{Z}^q$ iff we have $\mathrm{L}(\mathsf{P}_{\boldsymbol{Z}^q})\le 0$ for

$$\mathrm{L}(\mathsf{P}_{\boldsymbol{Z}^q}) = \min_{\substack{\mathsf{P}_1\\ \|\hat{\mathsf{P}}_1-\hat{\mathsf{P}}_0\|_2=\zeta}} \sum_{z\in\mathcal{Z}} \mathsf{P}_{\boldsymbol{Z}^q}[z]\log\frac{\mathsf{P}_0[z]}{\mathsf{P}_1[z]}.$$

Since $\|f\|_2^2=\frac{1}{d}\|\hat{f}\|_2^2$ for any function $f:\mathcal{Z}\to\mathbf{R}$, by writing $\mathsf{P}_0[z]=\frac{1}{d}$ and assuming that $\mathsf{P}_{\boldsymbol{Z}^q}[z]-\mathsf{P}_0[z]$ and $\mathsf{P}_1[z]-\mathsf{P}_0[z]$ are negligible to the first order, the above sum approximates to

$$\mathrm{L}(\mathsf{P}_{\boldsymbol{Z}^q}) \approx \frac{d}{\ln 2}\left(\frac{1}{2}\|\mathsf{P}_1-\mathsf{P}_0\|_2^2 - \max_{\mathsf{P}_1}\sum_{z\in\mathcal{Z}}\left(\mathsf{P}_{\boldsymbol{Z}^q}[z]-\frac{1}{d}\right)\left(\mathsf{P}_1[z]-\frac{1}{d}\right)\right)$$

which is clearly reached when $\mathsf{P}_1[z]-\frac{1}{d}$ is proportional to $\mathsf{P}_{\boldsymbol{Z}^q}[z]-\frac{1}{d}$. It is negative iff $\|\mathsf{P}_{\boldsymbol{Z}^q}-\mathsf{P}_0\|_2 \ge \frac{1}{2}\|\mathsf{P}_1-\mathsf{P}_0\|_2$. So the best distinguisher accepts $\boldsymbol{Z}^q$ iff $\|\mathsf{P}_{\boldsymbol{Z}^q}-\mathsf{P}_0\|_2 \ge \frac{\zeta}{2\sqrt{d}}$. We conclude by the following heuristic result.

Theorem 10. *If $Z=\chi(X)$ where χ is a character of order d, the best distinguisher between the null hypothesis that Z is uniformly distributed in*

the range of χ and the alternate hypothesis that Z follows some unknown distribution P_1 of known flatness $\zeta = \|\hat{\mathsf{P}}_1 - \hat{\mathsf{P}}_0\|_2$ is defined by

$$\mathsf{A}_q(\boldsymbol{Z}^q) = 1 \Leftrightarrow \sum_z \frac{\left(\mathsf{P}_{\boldsymbol{Z}^q}[z] - \frac{1}{d}\right)^2}{\frac{1}{d}} \geq \frac{\zeta^2}{4}.$$

It has an advantage approximated by

$$1 - \mathrm{Adv}_q \approx e^{-\frac{q}{8}\zeta^2}.$$

All in all, this is nothing but a χ^2 test on the frequencies with threshold $\frac{\zeta^2}{4}$.

5 Conclusion

We provided a precise asymptotic expression for the best distinguisher between two given distributions. We gave a simple approximation of this in terms of the Euclidean distance between the two distributions. We derived a link to the spectral analysis of distributions. We studied the problem of distinguishing one distribution from a set of distributions. This lead us to generalize linear cryptanalysis to arbitrary Abelian groups with order not necessarily equal to 2.

References

1. Thomas Baignères. *Quantitative Security of Block Ciphers: Design and Cryptanalysis Tools*. PhD Thesis, EPFL, (expected) 2008.
2. Thomas Baignères, Pascal Junod, and Serge Vaudenay. How far can we go beyond linear cryptanalysis? In P.J. Lee, editor, *Advances in Cryptology* - ASIACRYPT*'04*, volume 3329 of *LNCS*, pages 432–450. Springer-Verlag, 2004.
3. Thomas Baignères, Jacques Stern, and Serge Vaudenay. Linear cryptanalysis of non binary ciphers. In Carlisle M. Adams, Ali Miri, and Michael J. Wiener, editors, *Selected Areas in Cryptography - SAC 07*, volume 4876 of *LNCS*, pages 184–211. Springer-Verlag, 2007.
4. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons, 1991.
5. Mitsuru Matsui. Linear cryptanalysis method for DES cipher. In Tor Helleseth, editor, *Advances in Cryptology - EUROCRYPT'93: Workshop on the Theory and Application of Cryptographic Techniques, Lofthus, Norway, May 1993, Proceedings*, volume 765 of *LNCS*, pages 386–397. Springer-Verlag, May 1993.
6. Jerzy Neyman and Egon S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231:289–337, 1933.
7. Ivan N. Sanov. On the probability of large deviations of random variables. *Mat. Sbornik*, 42:11–44, 1957.