# Sublinear Bounds on the Distinguishing Advantage for Multiple Samples

Serge Fehr[1,2] and Serge Vaudenay[3]

[1] CWI, The Netherlands (serge.fehr@cwi.nl)
[2] Mathematical Institute, Leiden University, The Netherlands
[3] EPFL, Switzerland (serge.vaudenay@epfl.ch)

**Abstract.** The maximal achievable advantage of a (computationally unbounded) distinguisher to determine whether a source $Z$ is distributed according to distribution $P_0$ or $P_1$, when given access to *one* sample of $Z$, is characterized by the statistical distance $d(P_0, P_1)$. Here, we study the distinguishing advantage when given access to *several i.i.d. samples* of $Z$. For $n$ samples, the advantage is then naturally given by $d(P_0^{\otimes n}, P_1^{\otimes n})$, which can be bounded as $d(P_0^{\otimes n}, P_1^{\otimes n}) \leq n \cdot d(P_0, P_1)$. This bound is tight for some choices of $P_0$ and $P_1$; thus, *in general*, a linear increase in the distinguishing advantage is unavoidable.
In this work, we show new and improved bounds on $d(P_0^{\otimes n}, P_1^{\otimes n})$ that circumvent the above pessimistic observation. Our bounds assume, necessarily, certain additional information on $P_0$ and/or $P_1$ beyond, or instead of, a bound on $d(P_0, P_1)$; in return, the bounds grow as $\sqrt{n}$, rather than linearly in $n$. Thus, whenever applicable, our bounds show that the number of samples necessary to distinguish the two distributions is substantially larger than what the standard bound would suggest.
Such bounds have already been suggested in previous literature, but our new bounds are more general and (partly) stronger, and thus applicable to a larger class of instances.
In a second part, we extend our results to a modified setting, where the distinguisher only has *indirect* access to the source $Z$. By this we mean that instead of obtaining samples of $Z$, the distinguisher now obtains i.i.d. samples that are chosen according to a probability distribution that depends on the (one) value produced by the source $Z$.
Finally, we offer applications of our bounds to the area of cryptography. We show on a few examples from the cryptographic literature how our bounds give rise to improved results. For instance, importing our bounds into the analyses of Blondeau et al. for the security of block ciphers against multidimensional linear and truncated differential attacks, we obtain immediate improvements to their results.

## 1 Introduction

### 1.1 Motivation and Background

(In)distinguishability of probability distributions is a concept that is of fundamental importance in cryptography, for instance in the context of defining and

analyzing security of cryptographic schemes. It is well known that for a computationally unbounded distinguisher, given access to *one* sample of the source (i.e. random variable) $Z$ in question, the maximal achievable advantage in distinguishing whether $Z$ is distributed according to distribution $P_0$ or $P_1$, is given by the statistical distance

$$d(P_0, P_1) = \frac{1}{2} \sum_z |P_0(z) - P_1(z)| .$$

If the distinguisher has access to *multiple* samples $Z_1, \ldots, Z_n$ instead, each $Z_i$ being *identically and independently distributed* (i.i.d) according to $P_0$ or $P_1$, the best distinguishing advantage is then given by the statistical distance of the respective product distributions $P_b^{\otimes n}(z_1, \ldots, z_n) := P_b(z_1) \cdots P_b(z_n)$. Unfortunately, in general it is not easy to estimate $d(P_0^{\otimes n}, P_1^{\otimes n})$; a simple and commonly used bound is

$$d(P_0^{\otimes n}, P_1^{\otimes n}) \leq n \cdot d(P_0, P_1) . \tag{1}$$

This bound is very useful in that it is universally applicable, and in case no more information on $P_0$ and $P_1$ is available, it is the best one can hope for. E.g., if $P_0$ and $P_1$ are Boolean distributions with $P_0(1) = \epsilon$ and $P_1(1) = 0$, then $d(P_0, P_1) = \epsilon$ and $d(P_0^{\otimes n}, P_1^{\otimes n}) = 1 - (1 - \epsilon)^n \approx n\epsilon$ when $\epsilon \ll \frac{1}{n}$. Thus, the bound (1) is tight *in general*, if one has no additional information on $P_0$ and $P_1$.

However, in typical examples, one would expect to have some more information available on $P_0$ and $P_1$. In such cases, one may then hope for a better bound on the distinguishing advantage. Indeed, a few examples are known. For instance, Vaudenay [10] showed that for Boolean distributions and $P_1$ being the uniform distribution $U$, the bound

$$d(P_0^{\otimes n}, U^{\otimes n}) \leq 4\sqrt{n} \cdot d(P_0, U) \tag{2}$$

holds. A somewhat generalized variant of this is by Renner [8], which states that for any pair of (not necessarily Boolean) distributions, it holds that

$$d(P_0^{\otimes n}, P_1^{\otimes n}) \leq \sqrt{\frac{n}{2\bar{p}}} \cdot d(P_0, P_1) \tag{3}$$

where $\bar{p} := \min_z \min\{P_0(z), P_1(z)\}$ and the outer min is over all $z$ with $P_0(z) \neq P_1(z)$. Finally, one can also obtain a bound that grows with $\sqrt{n}$ by means of the *Rényi divergence* $D_\alpha$. Indeed, using basic properties of $D_\alpha$, and applying *Gilardoni's inequality* [5,3] if $0 < \alpha < 1$, and *Pinsker's inequality* if $\alpha = 1$, one immediately obtains the bound

$$d(P_0^{\otimes n}, P_1^{\otimes n}) \leq \sqrt{n} \cdot \sqrt{\frac{1}{2\alpha} D_\alpha(P_0 \| P_1)}$$

for $\alpha$ in the range $0 < \alpha \leq 1$. The case $\alpha = 1$, where the Rényi divergence is referred to as *Kullback-Leibler Divergence* (or KL divergence), was for instance used by Pöppelmann, Ducas, and Güneysu [6,7, Lemma 1].

In our work here, we give new support to this general hypothesis: we show new, non-trivial bounds on $d(P_0^{\otimes n}, P_1^{\otimes n})$ that apply if one has some more control over $P_0$ and $P_1$, beyond — or instead of — a bound on $d(P_0, P_1)$. Our bounds can be appreciated as generalizations and unification of (2) and (3) above.

Like the above examples, all our bounds, when applicable, show that the distinguishing advantage grows as $\sqrt{n}$ only, compared to the linear growth implied by (1). This means that in those cases, the number of samples necessary to distinguish the two distributions is substantially larger than what the bound (1) would suggest, i.e. quadratically more samples are actually needed. We discuss this on several concrete examples from the cryptographic literature, and we show how our new bounds give immediate rise to improved results.

Next to the concrete technical results and the applications discussed below, the goal of this work is to showcase the usefulness in the area of cryptography of the various notions of distance measures and their relations as studied in pure information theory (see e.g. [9] and the reference therein).

### 1.2 Our Technical Results

*New Bounds on the Distinguishing Advantage.* We show new bounds on the distinguishing advantage when given access to $n$ i.i.d. samples. Our first bound is a generalization (and slight improvement) of (2) that applies also for a non-uniform $P_1$; it shows that for any two Boolean distributions

$$d(P_0^{\otimes n}, P_1^{\otimes n}) \leq \sqrt{\frac{n}{2q(1-q)}}\, d(P_0, P_1) \tag{4}$$

where $q := P_1(0)$.

Our second bound removes the condition on the distributions being Boolean and is in terms of the 2-distance of $P_0$ and $P_1$. It states that

$$d(P_0^{\otimes n}, P_1^{\otimes n}) \leq \sqrt{\frac{n}{2\min_{z\in\Delta} P_1(z)}} \cdot \|P_0 - P_1\|_2\,, \tag{5}$$

where $\Delta = \{z \in \mathcal{Z} \;:\; P_0(z) \neq P_1(z)\}$. This improves upon Renner's bound (3) in that it requires control over the small probabilities of *one* of the two distributions only, and in that the 2-distance may be substantially smaller than the statistical distance. Also, compared to Renner's proof of (3), which relies on a cumbersome derivative estimation [8, Lemma 2.2] that goes over two pages, our proof is significantly simpler.

Using that $\|P_0 - P_1\|_2 \leq 2d(P_0, P_1)$, we thus obtain the bound

$$d(P_0^{\otimes n}, P_1^{\otimes n}) \leq \sqrt{\frac{2n}{\min_{z\in\Delta} P_1(z)}} \cdot d(P_0, P_1)\,, \tag{6}$$

which still requires control over the small probabilities of one of the two distributions only, but it is slightly worse in the constant factor than (3) .

Finally, applying (5) to $P_1 = U$, the uniform distribution over $\mathcal{Z}$, and setting $N = \#\mathcal{Z}$, we obtain the bound

$$d(P_0^{\otimes n}, U^{\otimes n}) \leq \sqrt{\frac{nN}{2}} \cdot \|P_0 - U\|_2 \leq \sqrt{2nN} \cdot d(P_0, U)\,, \tag{7}$$

which again can be appreciated as a generalization (and slight improvement) of Vaudenay's bound (2), now to non-Boolean distributions.

*Sources with Indirect Access.* We also study a variation of the setting discussed so far, where we considered a distinguisher with *direct access* to the source. Here, we show extensions of our results, i.e., bounds on the distinguishing advantage that grow as $\sqrt{n}$, when the distinguisher has *indirect* access only to the source.

Here, the distinguisher's goal is still to decide if a source $Z$ is distributed according to one or another distribution, which we now refer to as $Q_0$ and $Q_1$, but now he needs to do so by means of samples $W_1, \ldots, W_n$ that are obtained as follows: conditioned on that $Z = z$, which happens with probability either $Q_0(z)$ or $Q_1(z)$, the samples $W_i$ are i.i.d. according to a distribution $P_z$ *that depends on* $z$. Formally, given that $Z$ has distribution $Q_b$, the joint distribution of $W_1, \ldots, W_n$ is given by $\bar{P}_b^n := \sum_z Q_b(z)\, P_z^{\otimes n}$.

This setting naturally occurs in cryptography. E.g., let $Z$ be the key of a cipher following either a distribution $Q_0$ or a distribution $Q_1$. If we assume that the adversary obtains random plaintext/ciphertext pairs $W_i$ for that key, and lets say that the adversary wants to determine one bit of the secret (namely, whether it followed $Q_0$ or $Q_1$), then $d(\bar{P}_0^n, \bar{P}_1^n)$ offers a bound on the adversary's advantage.

Based on our bounds above, we show two bounds on $d(\bar{P}_0^n, \bar{P}_1^n)$, which both grow as $\sqrt{n}$. Like (4), (7) and (6) above, they differ in what kind of additional information is needed on the distributions $Q_0$ and $Q_1$, and on $\{P_z\}_{z \in \mathcal{Z}}$, for the bound to be meaningful. For the details, we refer to Section 4.

### 1.3 Applications

We show three immediate applications of our new bounds in the area of cryptography. The first two applications improve on the results of Blondeau et al. [1]. The first one is an improvement on the security bound for *multidimensional linear cryptanalysis* derived in [1], and the second application is an improvement on the security bound for *truncated differential cryptanalysis* derived in [1]. In both cases, our techniques apply very directly and enable to improve both bounds by a factor $\sqrt{n}$. Interestingly, such improvements were actually anticipated by Blondeau et al., but proving them was outside the scope of their techniques. We thus solve the open problems mentioned in [1].

The third application is to *decorrelation theory*, as introduced by Vaudenay [10]. By means of our bounds, we can improve the bound in [10] by a factor $\Theta(n^{1/6})$. This enables to prove the security of ciphers against iterated attacks for a larger number of iterations than was known before.

## 2 Preliminaries

### 2.1 The $p$-Norm and the Statistical Distance

Throughout the document, $\mathcal{Z}$ is a finite, non-empty set. We recall that for parameter $1 \leq p \leq \infty$, the *$p$-norm* of a function $f : \mathcal{Z} \to \mathbb{R}$ is defined as

$$\|f\|_p := \Big( \sum_{z \in \mathcal{Z}} |f(z)|^p \Big)^{1/p},$$

with the natural understanding that $\|f\|_\infty = \max_z |f(z)|$.

The definition of the $p$-norm obviously applies to *distributions* as well, i.e., to functions $P : \mathcal{Z} \to [0,1]$ with $\sum_z P(z) = \|P\|_1 = 1$. In particular, the *statistical distance* of two distributions $P_0, P_1 : \mathcal{Z} \to [0,1]$ is defined as

$$d(P_0, P_1) := \frac{1}{2}\|P_0 - P_1\|_1 = \frac{1}{2} \sum_z |P_0(z) - P_1(z)| \,.$$

We recall some basic properties of the statistical distance that are relevant for us. From the well-known fact that the statistical distance equals the *total variation distance*, i.e., $d(P_0, P_1) = \max_{E \subseteq \mathcal{Z}} \big|P_0(E) - P_1(E)\big|$, it follows that the statistical distance measures the maximal distinguishing advantage for (computationally unbounded) distinguishers that obtain one sample.

The statistical distance is *jointly convex*: if $P_0^i$ and $P_1^i$ are distributions that depend on some parameter $i \in \mathcal{I}$, and $\bar{P}_0 := \sum_i Q(i) P_0^i$ and $\bar{P}_1 := \sum_i Q(i) P_1^i$ are the corresponding mixtures with respect to distribution $Q : \mathcal{I} \to [0,1]$, then

$$d\big(\bar{P}_0, \bar{P}_1\big) \leq \sum_i Q(i)\, d(P_0^i, P_1^i) \,. \tag{8}$$

This follows immediately from basic properties of the 1-norm.

Finally, it is easy to see that $d(P_0 \otimes P', P_1 \otimes P') = d(P_0, P_1)$, where a *product distribution* $P \otimes P'$ is defined by $P \otimes P' : \mathcal{Z} \times \mathcal{Z}' \to [0,1]$, $(z, z') \mapsto P(z)P'(z')$. Together with the triangular inequality, this implies *subadditivity* for product distributions, i.e., $d(P_0 \otimes P_0', P_1 \otimes P_1') \leq d(P_0, P_1) + d(P_0', P_1')$. Thus, in particular, writing $P^{\otimes n}$ for $P \otimes \cdots \otimes P$ ($n$ times), we get inequality (1).

### 2.2 The Kullback-Leibler and the Rényi Divergence

The *Kullback-Leibler divergence* (or KL divergence) between two distributions $P_0, P_1 : \mathcal{Z} \to [0,1]$ is defined by

$$D_{\mathsf{KL}}(P_0 \| P_1) = \sum_{\substack{z \in \mathcal{Z} \\ P_0(z) > 0}} P_0(z) \ln \frac{P_0(z)}{P_1(z)} \,,$$

with the convention that $D_{\mathsf{KL}}(P_0 \| P_1) = \infty$ if the support of $P_0$ is not included in the support of $P_1$.

The Kullback-Leibler divergence is additive in the sense that

$$D_{\mathsf{KL}}(P_0 \otimes Q_0 \| P_1 \otimes Q_1) = D_{\mathsf{KL}}(P_0 \| P_1) + D_{\mathsf{KL}}(Q_0 \| Q_1)\,.$$

Consequently, we have

$$D_{\mathsf{KL}}(P_0^{\otimes n} \| P_1^{\otimes n}) = n \cdot D_{\mathsf{KL}}(P_0 \| P_1)\,.$$

It relates to the statistical distance as follows.

**Theorem 1 (Pinsker inequality).** *For any two distributions $P_0$ and $P_1$, we have*

$$d(P_0, P_1) \leq \sqrt{\frac{1}{2} D_{\mathsf{KL}}(P_0 \| P_1)}\,.$$

The Kullback-Leibler divergence is a special case of the *Rényi divergence*

$$D_\alpha(P_0 \| P_1) = \frac{1}{\alpha - 1} \ln \sum_{\substack{z \in \mathcal{Z} \\ P_0(z) > 0}} P_0(z)^\alpha P_1(z)^{1-\alpha}\,,$$

defined for any $0 < \alpha \neq 1$, where the Kullback-Leibler divergence is recovered in the limit $\alpha \to 1$. Like the Kullback-Leibler divergence, the Rényi divergence is additive, and the Pinsker inequality generalizes as follows.

**Theorem 2 (Gilardoni inequality [5,3]).** *For any two distributions $P_0$ and $P_1$ and $0 < \alpha < 1$, we have*

$$d(P_0, P_1) \leq \sqrt{\frac{1}{2\alpha} D_\alpha(P_0 \| P_1)}\,.$$

### 2.3 The Neyman Divergence

We also make use of the *Neyman $\chi^2$ divergence*, which we denote by $D_{\mathsf{N}}$. For distributions $P_0, P_1 : \mathcal{Z} \to [0, 1]$, it is defined as

$$D_{\mathsf{N}}(P_0 \| P_1) = \sum_{\substack{z \in \mathcal{Z} \\ P_1(z) > 0}} \frac{\big(P_0(z) - P_1(z)\big)^2}{P_1(z)}\,,$$

with the convention that $D_{\mathsf{N}}(P_0 \| P_1) = \infty$ if the support of $P_0$ is not included in the support of $P_1$.

**Theorem 3.** *For any two distributions $P_0$ and $P_1$, we have*

$$D_{\mathsf{KL}}(P_0 \| P_1) \leq D_{\mathsf{N}}(P_0 \| P_1)\,.$$

This was proven by Dai et al. [2] but has been known in the information-theory community for longer (see e.g. [9]); we recall here the proof for completeness.

*Proof.* Multiplying out the square in the enumerator, we obtain

$$D_{\mathsf{N}}(P_0\|P_1) = \sum_z \frac{P_0(z)^2}{P_1(z)} - 1 = e^{\ln\left(\sum_z \frac{P_0(z)^2}{P_1(z)}\right)} - 1$$
$$\geq \ln\left(\sum_z \frac{P_0(z)^2}{P_1(z)}\right) \geq \sum_z P_0(z) \ln \frac{P_0(z)}{P_1(z)} = D_{\mathsf{KL}}(P_0\|P_1)\,,$$

where the first inequality uses $e^x - 1 \geq x$, which is verified by having equality for $x = 0$ and comparing derivatives, and the second inequality is Jensen's inequality, exploiting concavity of ln. □

### 2.4 Warm-up Observation

We discuss yet another distance measure for distributions $P_0, P_1 : \mathcal{Z} \to [0,1]$. The *fidelity* of $P_0$ and $P_1$ (aka. *Bhattacharyya distance*) is defined as

$$F(P_0, P_1) := \sum_{z\in\mathcal{Z}} \sqrt{P_0(z)P_1(z)}\,.$$

Like the statistical distance, the fidelity of two distributions is always in the range 0 to 1; this follows immediately from the Chauchy-Schwarz inequality. We emphasize though that the fidelity is *not* a metric in the mathematical sense. In particular, $F(P_0, P_1)$ is *small* when $P_0$ and $P_1$ are *far apart*, and it is *large*, i.e. close to 1, when $P_0$ and $P_1$ are *close*. As a matter of fact, it turns out that $H(P_0, P_1) := \sqrt{1 - F(P_0, P_1)}$ is a metric, known as *Hellinger distance*. Nevertheless, it is useful to consider the fidelity directly as a measure of distance. Fidelity is related to the Rényi divergence of order $\frac{1}{2}$ by

$$D_{1/2}(P_0\|P_1) = -2\ln(F(P_0, P_1))\,.$$

Sometimes referred to as the *Fuchs – van de Graaf inequalities*, the following relates the fidelity to the statistical distance.

**Theorem 4.** *For distributions $P_0$ and $P_1$,*

$$1 - F(P_0, P_1) \leq d(P_0, P_1) \leq \sqrt{1 - F(P_0, P_1)^2}\,.$$

We conclude this brief introduction to the fidelity by pointing out that the fidelity is *multiplicative* for product distributions, and thus in particular

$$F(P_0^{\otimes n}, P_1^{\otimes n}) = F(P_0, P_1)^n\,. \tag{9}$$

We show here a simple standard application of the fidelity and its properties. A typical question is to wonder how many samples $n$ are needed to distinguish two known distributions $P_0$ and $P_1$ with constant advantage $t$. Motivated by this question, let $n_t$ be such that $n \geq n_t \iff d(P_0^{\otimes n}, P_1^{\otimes n}) \geq t$ hold for all $n$. Using (1), we only get the relatively crude bound $n_t \geq t/d(P_0, P_1)$. However, if

we actually control the fidelity $F(P_0, P_1)$ and set[4] $\epsilon := -2\log_2(F(P_0, P_1))$ then we obtain from Theorem 4 and property (9) that

$$-\frac{1}{\epsilon}\log_2(1-t^2) \leq n_t \leq -\frac{2}{\epsilon}\log_2(1-t)\,,$$

which is a much more precise estimate of the threshold number $n_t$. For instance, $n_{0.5}$ is in the range $0.41/\epsilon \leq n_{0.5} \leq 2/\epsilon$. In Appendix A, we work out a concrete application of this in the context of side-channel attacks.

## 3 Sublinear Bounds on the Statistical Distance

We present here our new bounds on the statistical distance of i.i.d. samples. At the core of the bounds is the following lemma.[5]

**Lemma 5.** *For any two distributions $P_0, P_1 : \{0,1\} \to [0,1]$ and any integer $n$,*

$$d(P_0^{\otimes n}, P_1^{\otimes n}) \leq \sqrt{\frac{n}{2} D_{\mathsf{N}}(P_0 \| P_1)}\,.$$

*Proof.* We have

$$d(P_0^{\otimes n}, P_1^{\otimes n}) \leq \sqrt{\frac{1}{2} D_{\mathsf{KL}}(P_0^{\otimes n} \| P_1^{\otimes n})} = \sqrt{\frac{n}{2} D_{\mathsf{KL}}(P_0 \| P_1)} \leq \sqrt{\frac{n}{2} D_{\mathsf{N}}(P_0 \| P_1)}\,,$$

where the first inequality is Pinsker's inequality, the equality is by the additivity of the Kullback-Leibler divergence, and the final inequality is by Theorem 3. □

### 3.1 A Bound for Boolean Distributions

We first consider Boolean distributions.

**Theorem 6.** *Let $P_0, P_1 : \{0,1\} \to [0,1]$ be two Boolean distributions. Then*

$$d(P_0^{\otimes n}, P_1^{\otimes n}) \leq \sqrt{\frac{n}{2q(1-q)}}\, d(P_0, P_1)\,,$$

*where $q := P_1(0)$.*

We observe that in the special case of $P_1 = U$, the uniform distribution on $\{0,1\}$, we recover a slightly improved version of the bound (2) from [10].

*Proof.* Setting $p := P_0(0)$ and $q := P_1(0)$, we have

$$D_{\mathsf{N}}(P_0 \| P_1) = \frac{(p-q)^2}{q} + \frac{(q-p)^2}{1-q} = \frac{(p-q)^2}{q(1-q)} = \frac{d(P_0, P_1)^2}{q(1-q)}\,.$$

The claim thus holds by Lemma 5. □

[4] As a matter of fact, $\epsilon$ is the Rényi divergence measured in bits.

[5] This lemma was hinted at by an anonymous reviewer. It improves and simplifies on an earlier version of this paper.

### 3.2 Bounds for Non-Boolean Distributions

Here, we drop the assumption on the distributions $P_0$ and $P_1$ being Boolean but instead assume that we have control over the small probabilities of one of the two distributions. Concretely, we assume control over

$$\min_{\Delta}(P_1) := \min_{z \in \Delta} P_1(z)$$

where $\Delta := \Delta(P_0, P_1) := \{z \in \mathcal{Z} : P_0(z) \neq P_1(z)\}$. This is well defined for $P_0 \neq P_1$.

**Theorem 7.** *For different distributions $P_0, P_1 : \mathcal{Z} \to [0,1]$ and $\min_{\Delta}(P_1)$ as above,*

$$d(P_0^{\otimes n}, P_1^{\otimes n}) \leq \sqrt{\frac{n}{2\min_{\Delta}(P_1)}} \cdot \|P_0 - P_1\|_2 \,.$$

*Proof.* If there exists $z \in \mathcal{Z}$ with $P_0(z) \neq 0 = P_1(z)$ then $\min_{\Delta}(P_1) = 0$ and the claim holds trivially, with the right hand side being $\infty$ then. If no such $z \in \mathcal{Z}$ exists then $D_{\mathsf{N}}(P_0 \| P_1)$ is finite and equal to

$$\begin{aligned} D_{\mathsf{N}}(P_0\|P_1) &= \sum_{\substack{z \in \mathcal{Z} \\ P_1(z) > 0}} \frac{\big(P_0(z) - P_1(z)\big)^2}{P_1(z)} = \sum_{\substack{z \in \mathcal{Z} \\ P_0(z) \neq P_1(z) > 0}} \frac{\big(P_0(z) - P_1(z)\big)^2}{P_1(z)} \\ &\leq \sum_{z \in \mathcal{Z}} \frac{\big(P_0(z) - P_1(z)\big)^2}{\min_{\Delta}(P_1)} = \frac{\|P_0 - P_1\|_2^2}{\min_{\Delta}(P_1)} \,. \end{aligned}$$

The claimed bound then holds by Lemma 5. □

Recalling that $\|P_0 - P_1\|_2 \leq 2d(P_0, P_1)$, we obtain the following.

**Corollary 8.** *For different distributions $P_0, P_1 : \mathcal{Z} \to [0,1]$ and $\min_{\Delta}(P_1)$ as above,*

$$d(P_0^{\otimes n}, P_1^{\otimes n}) \leq \sqrt{\frac{2n}{\min_{\Delta}(P_1)}} \cdot d(P_0, P_1) \,.$$

This bound can be appreciated as a variant of Renner's bound (3), which we rephrase here as

$$d(P_0^{\otimes n}, P_1^{\otimes n}) \leq \sqrt{\frac{n}{2\min_{\Delta}(P_0 \cup P_1)}} \cdot d(P_0, P_1)$$

for

$$\min_{\Delta}(P_0 \cup P_1) := \min_{z \in \Delta} \min\big\{P_0(z), P_1(z)\big\} \,,$$

where $\Delta$ is as above. Our new bound improves on this in that it requires control over the small probabilities of *one* of the two distributions only (but is slightly worse in the constant factor). Additionally, compared to Renner's result [8], we also have a substantially simpler proof.

Applying Theorem 7 to $P_1$ being the uniform distribution with $U(z) = 1/|\mathcal{Z}|$ for all $z \in \mathcal{Z}$, we obtain the following.

**Corollary 9.** *Let $P_0$ be a distribution over a set $\mathcal{Z}$ with cardinality $N := |\mathcal{Z}|$, and let $U$ be the uniform distribution over the same set $\mathcal{Z}$. Then, we have*

$$d(P_0^{\otimes n}, U^{\otimes n}) \leq \sqrt{\frac{nN}{2}} \cdot \|P_0 - U\|_2 \leq \sqrt{2nN} \cdot d(P_0, U)\,.$$

This as well can be appreciated as a generalization (and slight improvement) of the bound (2), now to non-Boolean distributions.

## 4 Indistinguishability for Sources with Indirect Access

The above bounds — on distinguishing whether a source $Z$ is distributed according to $P_0$ or $P_1$ — apply when the distinguisher has access to independently generated samples of $Z$. Here, we consider a variation of this problem where the distinguisher does not have direct access to the source $Z$, distributed according to $Q_0$ or $Q_1$; instead, the distinguisher obtains independent samples $W_1, \ldots, W_n$ of a source $W$ that depends on $Z$. Formally, conditioned on the event $Z = z$, which happens with probability $Q_0(z)$ or $Q_1(z)$, the joint distribution of $W_1, \ldots, W_n$ is given by $P_z^{\otimes n}$, where $\{P_z\}_{z \in \mathcal{Z}}$ is a given family of distributions over a set $\mathcal{W}$. Algorithmically, for a fixed guessing function $f$, the task of distinguishing $Q_0$ from $Q_1$ in this setting can be captured as illustrated in Figure 1, and the maximal distinguishing advantage is given by $d(\bar{P}_0^n, \bar{P}_1^n)$, where $\bar{P}_b^n := \sum_z Q_b(z) P_z^{\otimes n}$.

**Distinguisher**:
1: pick $z$ following $Q_b$
2: **for** $i = 1$ to $n$ **do**
3: pick $w_i$ following $P_z$
4: **end for**
5: output $f(w_1, \ldots, w_n)$

**Fig. 1.** Distinguisher with indirect access to $Z$.

Below and for the remainder, we use the following notation. Assuming the range $\mathcal{Z}$ of $Z$ to be in $\mathbb{R}$, and given that $Z$ is distributed according to $Q$ (which will either be $Q_0$ or $Q_1$), we let $E(Z)$ be the *expectation*, i.e., $E(Z) := \sum_z Q(z)\, z$, and correspondingly $E(g(Z)) := \sum_z Q(z)\, g(z)$ for any function $g$ on $\mathbb{R}$. Similarly, $V(Z)$ denotes the *variance* $V(Z) := E\big((Z - E(Z))^2\big) = E(Z^2) - E(Z)^2$. If we want to make $Q$ explicit, we write $E_Q(Z)$, $E_Q(g(Z))$ and $V_Q(Z)$ instead. This notation generalizes to $Z$ with range $\mathcal{Z}$ in an arbitrary vector space (with a given 2-norm): $E$ is obvious, and $V$ then becomes $V(Z) := E\big(\|Z - E(Z)\|_2^2\big)$.

### 4.1 The Boolean Case

We first consider the case of Boolean samples, i.e., where the $W_i$'s have values in $\{0, 1\}$. Since $P_z$ is fully defined by $P_z(1)$ then, we may assume without loss

of generality that $\mathcal{Z} \subset [0,1] \subset \mathbb{R}$ and $P_z(1) = z$ for any $z \in \mathcal{Z}$. In other words, the output of the source not only determines but *equals* the probability of each sample to be 1.

For $n = 1$ and guessing function $f(w_1) = w_1$, the distinguishing advantage is given by $E_{Q_0}(Z) - E_{Q_1}(Z)$. For $n = 2$ and $f(w_1, w_2) = w_1 w_2$, the advantage is

$$E_{Q_0}(Z^2) - E_{Q_1}(Z^2) = V_{Q_0}(Z) - V_{Q_1}(Z) + E_{Q_0}(Z)^2 - E_{Q_1}(Z)^2 \,.$$

Therefore, to keep the advantage low for arbitrary distinguishers, we obviously need $E_{Q_0}(Z) \approx E_{Q_1}(Z)$ (due to the $n = 1$ case) and $V_{Q_0}(Z) \approx V_{Q_1}(Z)$ (due to the $n = 2$ case). These two conditions are necessary. We show below that these conditions, and the assumption that the two variances are small, are also *sufficient* for indistinguishability.

**Theorem 10.** *Let $Q_0$ and $Q_1$ be two distributions for $Z$, and let $\{P_z\}_{z\in\mathcal{Z}}$ be a family of Boolean distributions. Then, for any $n$ and for $\bar{P}_0^n$ and $\bar{P}_1^n$ defined as above,*

$$d(\bar{P}_0^n, \bar{P}_1^n) \le \sqrt{\frac{n}{2}} \cdot \frac{\sqrt{V_{Q_0}(Z) + (E_{Q_0}(Z) - E_{Q_1}(Z))^2} + \sqrt{V_{Q_1}(Z)}}{\sqrt{E_{Q_1}(Z)\big(1 - E_{Q_1}(Z)\big)}}$$

We will show in Section 5.3 a direct application of this to [10].

*Proof.* Recall that the source $Z$ takes values $0 \le z \le 1$, and for any such $z$, $P_z$ is a Boolean distribution with $P_z(1) = z$. For an arbitrary but fixed $0 \le \mu \le 1$, we also consider the Boolean distribution $P_\mu$ with $P_\mu(1) = \mu$. Applying Theorem 6, we obtain

$$d(P_z^{\otimes n}, P_\mu^{\otimes n}) \le \sqrt{\frac{n}{2\mu(1-\mu)}}\, d(P_z, P_\mu) = \sqrt{n\frac{(z-\mu)^2}{2\mu(1-\mu)}} \,.$$

By convexity (8) of the statistical distance, and applying Jensen inequality with $\sqrt{\cdot}$ and using that $E((Z-\mu)^2) = V(Z) + (E(Z) - \mu)^2$, we obtain

$$d(\bar{P}_b^n, P_\mu^{\otimes n}) \le E_{Q_b}\big(d(P_Z^{\otimes n}, P_\mu^{\otimes n})\big) \le \sqrt{n\frac{V_{Q_b}(Z) + (E_{Q_b}(Z) - \mu)^2}{2\mu(1-\mu)}}$$

for $b \in \{0,1\}$. Hence, by triangular inequality,

$$d(\bar{P}_0^n, \bar{P}_1^n) \le \sqrt{n\frac{V_{Q_0}(Z) + (E_{Q_0}(Z) - \mu)^2}{2\mu(1-\mu)}} + \sqrt{n\frac{V_{Q_1}(Z) + (E_{Q_1}(Z) - \mu)^2}{2\mu(1-\mu)}} \,.$$

This holds for any $\mu$. We can apply it to $\mu = E_{Q_1}(Z)$ and obtain the result. □

### 4.2 The Non-Boolean Case

Here, we consider the non-Boolean case where the samples $W_i$ have range $\mathcal{W}$ of arbitrary size $N$. Our result gives a meaningful bound if the distributions $P_z$ are close to uniform on average.

**Theorem 11.** *Let $Q_0$ and $Q_1$ be two distributions for $Z$, and let $\{P_z\}_{z\in\mathcal{Z}}$ be a family of distributions over a set of size $N$. Then, for any $n$ and for $\bar{P}_0^n$ and $\bar{P}_1^n$ defined as above,*

$$d(\bar{P}_0^n, \bar{P}_1^n) \le \sqrt{\frac{nN}{2}} \cdot \Big( E_{Q_0}\big(\|P_Z - U\|_2\big) + E_{Q_1}\big(\|P_Z - U\|_2\big)\Big) .$$

For clarification, given that $Z$ is a random variable, we recall that $P_Z$ is a random variable as well; its range being $\{P_z \ : \ z \in \mathcal{Z}\}$, a subset of the distributions over $\mathcal{W}$. As such, $\|P_Z - U\|_2$ is then a real-valued random variable, and so its expectation is well defined.

*Proof.* Applying Corollary 9, we have $d(P_z^{\otimes n}, U^{\otimes n}) \le \sqrt{\frac{nN}{2}} \cdot \|P_z - U\|_2$ , and by convexity (8) of the statistical distance, we then obtain

$$d(\bar{P}_b^n, U^{\otimes n}) \le E_{Q_b}\big(d(P_Z^{\otimes n}, U^{\otimes n})\big) \le \sqrt{\frac{nN}{2}} \cdot E_{Q_b}\left(\|P_Z - U\|_2\right) .$$

The claim then follows from triangular inequality. □

## 5 Applications

We discuss three direct applications of our new bounds in the context of block-cipher security, giving rise to immediate improvements to results in [1] and [10].

### 5.1 Resistance to Multidimensional Linear Cryptanalysis

As a first application, we improve the security bound from Blondeau et al. [1] against *multidimensional linear* (ML) cryptanalysis. As considered in [1], for a block cipher $\mathsf{Enc}$ over $\ell$-bit blocks and a vector subspace $V$ of $\{0,1\}^\ell \times \{0,1\}^\ell$ spanned by a basis $(\alpha_1, \beta_1), \ldots, (\alpha_k, \beta_k)$, the so-called *linear masks*, an *ML distinguisher* works as described in Fig. 2. We refer to [1] for the motivation and for additional explanations.

Adopting the notation from Blondeau et al. [1], we write $p^{\mathsf{ML}}_{\mathsf{Enc}}$ for the probability that the distinguisher outputs 1 using $\mathsf{Enc}$ with an arbitrary but *fixed* key, and we write $p^{\mathsf{ML}}_{C_K}$ for the same probability but now considered as a random variable with the randomness stemming from the random choice of the key. We note that for $\mathsf{Enc}$ with a fixed key, the $b_i$'s are i.i.d. over $\{0,1\}^k$, and we denote the distribution of one $b_i$ by $D_{\mathsf{Enc}}$ then.

**Distinguisher ML** with oracle $\mathsf{Enc}$:

1: **for** $i = 1$ to $n$ **do**
2: pick a random $x \in \{0,1\}^\ell$
3: set $y = \mathsf{Enc}(x)$
4: **for** $j = 1$ to $k$ **do**
5: set $b_{i,j} = (\alpha_j \cdot x) \oplus (\beta_j \cdot y)$
6: **end for**
7: set $b_i = (b_{i,1}, \ldots, b_{i,k})$
8: **end for**
9: output $f(b_1, \ldots, b_n)$

**Fig. 2.** ML distinguisher

In [1, Thm. 19], it is shown that[6]

$$\left|E(p^{\mathsf{ML}}_{C_K}) - E(p^{\mathsf{ML}}_{C^*})\right| \le n\sqrt{2^k}\sqrt{2^{-\ell} + \frac{1}{4}\left\|[C_K]^2 - [C^*]^2\right\|_\infty} \tag{10}$$

where $\|[C_K]^2 - [C^*]^2\|_\infty$ denotes the *decorrelation of order 2*, i.e., twice the best non-adaptive advantage to distinguish the cipher $C_K$ from a uniformly random permutation $C^*$ using two queries. Our results allow us to improve this bound by a factor $\sqrt{2n}$.

At the core of the proof in [1] is the observation that for any two fixed encryption functions $\mathsf{Enc}$ and $\mathsf{Enc}^*$(with fixed keys), by [1, Lemma 14] and triangular inequality,

$$\left|p^{\mathsf{ML}}_{\mathsf{Enc}} - p^{\mathsf{ML}}_{\mathsf{Enc}^*}\right| \le \frac{n}{2}\sqrt{2^k}\|D_{\mathsf{Enc}} - D_{\mathsf{Enc}^*}\|_2 \le \frac{n}{2}\sqrt{2^k}\left(\|D_{\mathsf{Enc}} - U\|_2 + \|D_{\mathsf{Enc}^*} - U\|_2\right),$$

and then these 2-norms are further worked out, using [1, Lemma 12] etc. How exactly the bound (10) is then derived in [1] is not so important here; what is important is that the factor $n$ from above directly carries into (10).

Our improvement is now to apply triangular inequality to $d(D_{\mathsf{Enc}}, D_{\mathsf{Enc}^*})$ and then invoke Corollary 9 to show that

$$\left|p^{\mathsf{ML}}_{\mathsf{Enc}} - p^{\mathsf{ML}}_{\mathsf{Enc}^*}\right| \le \sqrt{n2^{k-1}}\left(\|D_{\mathsf{Enc}} - U\|_2 + \|D_{\mathsf{Enc}^*} - U\|_2\right)$$

instead. Then, we can argue exactly as in [1] to conclude the following, which in particular solves one of the open questions posed in [1].

**Theorem 12.** *For every cipher $C_K$ over $\ell$-bit blocks, the ML-advantage for $k$-linear masks, i.e., the advantage of an ML distinguisher as in Fig. 2, is bounded by*

$$\left|E(p^{\mathsf{ML}}_{C_K}) - E(p^{\mathsf{ML}}_{C^*})\right| \le \sqrt{n2^{k+1}}\sqrt{2^{-\ell} + \frac{1}{4}\left\|[C_K]^2 - [C^*]^2\right\|_\infty}\,.$$

[6] We point out that in the derivation of their bound, [1] uses $\sqrt{a+b}+\sqrt{a} \le 2\sqrt{a+b}$. If, instead, we use $\sqrt{a+b}+\sqrt{a} \le 2\sqrt{a+b/2}$, which hold by Jensen's inequality, we obtain the slightly improved version stated here in (10), with a factor $1/4$ instead $1/2$.

Alternatively, and slightly more directly, we can use Theorem 11 to argue that

$$\left|E(p^{\mathsf{ML}}_{C_K}) - E(p^{\mathsf{ML}}_{C^*})\right| \le \sqrt{n2^{k-1}} \cdot \Big(E_{Q_0}\big(\|P_Z - U\|_2\big) + E_{Q_1}\big(\|P_Z - U\|_2\big)\Big)\,,$$

and then apply Lemma 12, Theorem 1 and Lemma 17 and 18 from [1] to obtain the above bound.

### 5.2 Resistance to Truncated Differential Attacks

Here, we revisit and improve the security bound from Blondeau et al. [1] against what is knows as *truncated differential* (TD) attacks. Following the definition given in [1], for a block cipher $\mathsf{Enc}$ over $\ell$-bit blocks and for a vector space $V^\perp = V^\perp_{\mathsf{in}} \times V^\perp_{\mathsf{out}}$ with $V^\perp_{\mathsf{in}}$ having dimension $\ell - s > 0$, a TD distinguisher works as described in Fig. 3 to the right.

**Distinguisher TD** with oracle $\mathsf{Enc}$:
1: **for** $i = 1$ to $n$ **do**
2: pick $(x, x') \in (\{0,1\}^\ell)^2$ uniformly such that $x \oplus x' \in V^\perp_{\mathsf{in}}$
3: set $y = \mathsf{Enc}(x)$ and $y' = \mathsf{Enc}(x')$
4: set $b_i = 1_{((x,y)\oplus(x',y'))\in V^\perp}$
5: **end for**
6: output $f(b_1, \ldots, b_n)$

**Fig. 3.** TD distinguisher

Using notation similar to as above, [1, Thm. 21] shows that

$$\left|E(p^{\mathsf{TD}}_{C_K}) - E(p^{\mathsf{TD}}_{C^*})\right| \le n2^s\Big(2 \cdot 2^{-\ell} + \frac{1}{2}\big\|[C_K]^2 - [C^*]^2\big\|_\infty\Big)\,. \tag{11}$$

At the core of the proof of (11) is the bound

$$\begin{aligned}\left|p^{\mathsf{TD}}_{\mathsf{Enc}} - p^{\mathsf{TD}}_{\mathsf{Enc}^*}\right| &\le n\, d(D_{\mathsf{Enc}}, D_{\mathsf{Enc}^*}) \\ &= n2^s\left|p^{\mathsf{STD}}_{\mathsf{Enc}} - p^{\mathsf{STD}}_{\mathsf{Enc}^*}\right| \\ &\le n2^s\big(\left|p^{\mathsf{STD}}_{\mathsf{Enc}} - 2^{-\ell}\right| + \left|p^{\mathsf{STD}}_{\mathsf{Enc}^*} - 2^{-\ell}\right|\big)\end{aligned}$$

where $p^{\mathsf{STD}}_{\mathsf{Enc}}$ is defined in [1] and happens to coincide with $2^{-s}D_{\mathsf{Enc}}(1)$. Again, it is not relevant here how (11) is then derived from this; important for us is that the factor $n2^s$ carries into the bound.

Our improvement is obtained by applying Corollary 8 to the Boolean distributions $D_{\mathsf{Enc}}$ and $D_{\mathsf{ref}}$, with the latter defined as $D_{\mathsf{ref}}(1) = 2^{s-\ell}$. This gives

$$\begin{aligned} d(D_{\mathsf{Enc}}^{\otimes}, D_{\mathsf{ref}}^{\otimes}) &\leq \sqrt{\frac{n}{2\min(2^{s-\ell}, 1-2^{s-\ell})}}\, d(D_{\mathsf{Enc}}, D_{\mathsf{ref}}) \\ &= \sqrt{\frac{n}{2^{s-\ell+1}}}\, \left|2^s p_{\mathsf{Enc}}^{\mathsf{STD}} - 2^{s-\ell}\right| \\ &= \sqrt{n2^{s+\ell-1}} \left|p_{\mathsf{Enc}}^{\mathsf{STD}} - 2^{-\ell}\right| . \end{aligned}$$

Therefore, by triangular inequality, we obtain

$$\left|p_{\mathsf{Enc}}^{\mathsf{TD}} - p_{\mathsf{Enc}^*}^{\mathsf{TD}}\right| \leq d(D_{\mathsf{Enc}}^{\otimes}, D_{\mathsf{Enc}^*}^{\otimes}) \leq \sqrt{n2^{s+\ell-1}} \left(\left|p_{\mathsf{Enc}}^{\mathsf{STD}} - 2^{-\ell}\right| + \left|p_{\mathsf{Enc}^*}^{\mathsf{STD}} - 2^{-k}\right|\right) .$$

By the techniques of [1], this then result in the following bound, which improves (11) by a factor $\sqrt{n}\, 2^{(s-\ell+1)/2}$ and solves the second open problem from [1].

**Theorem 13.** *For every cipher $C_K$ over $\ell$-bit blocks, the TD-advantage for $V_{\mathsf{in}}^{\perp}$ of dimension $\ell - s$, i.e., the advantage of a TD distinguisher as in Fig. 3, is bounded by*

$$\left|E(p_{C_K}^{\mathsf{TD}}) - E(p_{C^*}^{\mathsf{TD}})\right| \leq \sqrt{n2^{s+\ell-1}} \left(2 \cdot 2^{-\ell} + \frac{1}{2} \left\| [C_K]^2 - [C^*]^2 \right\|_{\infty}\right) .$$

### 5.3 Decorrelation

As a last application, we now move to decorrelation theory, as introduced by Vaudenay [10]. Again, we consider a block cipher $\mathsf{Enc}$ over $\ell$-bit blocks.[7] As considered and studied in [10], an iterated distinguisher of order $q$ is a distinguisher as described in Figure 4 to the right, where $n$ is a positive integer, $D$ is a probability distribution over $(\{0,1\}^{\ell})^q$, and $T$ and $f$ are functions $T : (\{0,1\}^{\ell})^{2q} \to \{0,1\}$ and $f : \{0,1\}^n \to \{0,1\}$.

Expressed in terms of notation similar to as above, it was shown in [10, Th. 18] that the advantage of any such distinguisher with Boolean $T$ is bounded by

$$\left|E(p_{C_K}^{\mathsf{Iter}}) - E(p_{C^*}^{\mathsf{Iter}})\right| \leq 5 \sqrt[3]{n^2 \left(2\delta + \frac{5q^2}{2 \cdot 2^{\ell}} + \frac{3}{2}\epsilon\,,\right)} + n\epsilon \tag{12}$$

where $\epsilon := \left\| [C_K]^{2q} - [C^*]^{2q} \right\|_{\infty}$, i.e., twice the best advantage of a distinguisher making $2q$ non-adaptive queries, and $\delta$ is the probability that two given iterations (say for $i = 1$ and $i = 2$) would select at least one $x_j$ in common (not necessarily for the same index $j$). We let $Z$ be the probability that $b_i = 1$ in the probability space induced by $D$ with a fixed $\mathsf{Enc}$ and write $E(Z)$ for the expected value of $Z$

[7] As a matter of fact, here $\mathsf{Enc}$ may also be a pseudorandom functions (PRF), but we ignore this here and keep the notation consistent with above.

**Distinguisher Iter** with oracle $\mathsf{Enc}$:
1: **for** $i = 1$ to $n$ **do**
2: pick $(x_1, \ldots, x_q) \in (\{0,1\}^\ell)^q$ following distribution $D$
3: set $y_j = \mathsf{Enc}(x_j)$ for $j = 1, \ldots, q$
4: set $b_i = T(x_1, \ldots, x_q, y_1, \ldots, y_q) \in \{0,1\}$
5: **end for**
6: output $f(b_1, \ldots, b_n)$

**Fig. 4.** Iterated distinguisher

with respect to the corresponding distribution of a random $\mathsf{Enc}$. By definition, it holds that

$$|E_{C_K}(Z) - E_{C^*}(Z)| \le \frac{1}{2}\big\|[C_K]^q - [C^*]^q\big\|_\infty \le \frac{\epsilon}{2}\,.$$

Furthermore, by considering a distinguisher that samples $b_i$ and $b_i'$ and returns $b_i b_i'$, which is 1 with probability $Z^2$, we see that

$$|E_{C_K}(Z^2) - E_{C^*}(Z^2)| \le \frac{1}{2}\big\|[C_K]^{2q} - [C^*]^{2q}\big\|_\infty = \frac{\epsilon}{2}\,.$$

Given that $E_{C^*}(Z)^2 - E_{C_K}(Z)^2 = \big(E_{C^*}(Z) - E_{C_K}(Z)\big)\big(E_{C^*}(Z) + E_{C_K}(Z)\big) \le \epsilon$, it is then easy to see that

$$V_{C_K}(Z) - V_{C^*}(Z) = E_{C_K}(Z^2) - E_{C^*}(Z^2) + E_{C^*}(Z)^2 - E_{C_K}(Z)^2 \le \frac{3}{2}\epsilon\,.$$

Finally, it was shown in [10] that

$$V_{C^*}(Z) \le \delta + \frac{q^2}{4\cdot 2^\ell} + \frac{q^2}{2\cdot(2^\ell - q)}\,.$$

So, applying Theorem 10 we obtain the bound

$$\begin{aligned}
\big|E(p^{\mathsf{Iter}}_{C_K}) - E(p^{\mathsf{Iter}}_{C^*})\big| &\le \sqrt{\frac{n}{2}} \cdot \frac{\sqrt{V_{C_K}(Z) + (E_{C_K}(Z) - E_{C^*}(Z))^2} + \sqrt{V_{C^*}(Z)}}{\sqrt{E_{C^*}(Z)\big(1 - E_{C^*}(Z)\big)}} \\
&\le \sqrt{\frac{n}{2E_{C^*}(Z)(1 - E_{C^*}(Z))}}\left(\sqrt{\frac{3}{2}\epsilon + V_{C^*}(Z) + \epsilon^2} + \sqrt{V_{C^*}(Z)}\right) \\
&\le \sqrt{\frac{2n}{E_{C^*}(Z)(1 - E_{C^*}(Z))}}\sqrt{\delta + \frac{q^2}{4\cdot 2^\ell} + \frac{q^2}{2\cdot(2^\ell - q)} + \frac{5}{2}\epsilon}\,.
\end{aligned}$$

Thus, in summary, and using the notation from above, we obtain the following improved version of (12).

**Theorem 14.** *Let $C_K$ be a cipher over $\ell$-bit blocks. Then, for any positive integers $q$ and $n$, any distribution $D$ over $(\{0,1\}^\ell)^q$, and any functions $T$ :*

$(\{0,1\}^\ell)^{2q} \to \{0,1\}$ *and* $f : \{0,1\}^n \to \{0,1\}$*, the iterated distinguisher of order* $q$ *from Fig. 4 has an advantage*

$$\left|E(p^{\mathsf{Iter}}_{C_K}) - E(p^{\mathsf{Iter}}_{C^*})\right| \leq \sqrt{\frac{2n}{p(1-p)}}\sqrt{\delta + \frac{q^2}{4\cdot 2^\ell} + \frac{q^2}{2\cdot(2^\ell - q)} + \frac{5}{2}\epsilon}\,,$$

*where* $p := E_{C^*}(Z)$ *is the probability that* $b_i = 1$ *in case of a random permutation, and* $\epsilon$ *and* $\delta$ *are as in [10], i.e.,* $\epsilon = \|[C_K]^{2q} - [C^*]^{2q}\|_\infty$ *and* $\delta$ *is the probability that two given iterations have one* $x_j$ *in common (not necessarily for the same index* $j$*).*

This is better than (12) by a ratio which is asymptotically $\Theta(n^{1/6})$, but requires that we know $E_{C^*}(Z)$. This is normally the case though. We can further see that for $E_{C^*}(Z)$ close to $1/2$, we obtain security for $n \ll 1/\epsilon$ while the previous result (12) offers security only for $n \ll 1/\sqrt{\epsilon}$. We thus obtain security for significantly larger $n$.

## 6 Conclusion

We derived new bounds on the statistical distance $d(P_0^{\otimes n}, P_1^{\otimes n})$ of i.i.d. samples of a source $Z$ that is distributed according to distribution $P_0$ or $P_1$, as well as on the statistical distance of i.i.d. samples that are chosen according a distribution that depends on $Z$. All the bounds grow as $\sqrt{n}$ in the number $n$ of samples, and they are applicable if some additional information on the distributions is known.

We expect these new bounds to become useful tools in cryptography. Indeed, we demonstrated the usefulness on several applications in the context of block-cipher analysis. In all these examples, our bounds lead to an immediate improvement over the prior results.

## A Computing the Threshold Number in Power-Analysis Attacks

At the core of a power-analysis attack is the task to distinguish whether a source $X$ over a finite set $\mathcal{X}$ is distributed according to $P_0$ or $P_1$ when given i.i.d. samples of the form $X + N$, where $N$ is noise that follows a normal distribution with expected value 0 and standard deviation $\sigma$. Formally, given that $X$ is distributed according to $P_b$, the random variable $X + N$ has density

$$f_{Q_b}(t) = \sum_{x\in\mathcal{Z}} P_b(x)\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(t-x)^2}{2\sigma^2}}\,.$$

Hence, not worrying that we are now dealing with continuous random variables, we have

$$\begin{aligned} F(Q_0\|Q_1) &= \int_{-\infty}^{+\infty} \sqrt{f_{Q_0}(t) f_{Q_1}(t)}\, dt \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sqrt{\sum_{x,y\in\mathcal{X}} P_0(x) P_1(y) e^{-\frac{(t-x)^2+(t-y)^2}{2\sigma^2}}}\, dt\,. \end{aligned}$$

Thus, the threshold number of samples is given by

$$n_{1/2} = \frac{\theta}{-2\log_2\left(\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sqrt{\sum_{x,y\in\mathcal{X}} P_0(x) P_1(y) e^{-\frac{(t-x)^2+(t-y)^2}{2\sigma^2}}}\, dt\right)}$$

for some $0.41 \leq \theta \leq 2$.

As a concrete example, if $P_0$ and $P_1$ are such that $P_0(x) = 1 = P_1(y)$ for some fixed values $x, y \in \mathcal{X}$, then we get

$$\begin{aligned} F(Q_0\|Q_1) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sqrt{e^{-\frac{(t-x)^2+(t-y)^2}{2\sigma^2}}}\, dt \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(t-\frac{x+y}{2})^2+(\frac{x-y}{2})^2}{2\sigma^2}}\, dt \\ &= e^{-\frac{(x-y)^2}{8\sigma^2}} \end{aligned}$$

and thus

$$-2\log F(Q_0\|Q_1) = \frac{(x-y)^2}{(4\ln 2)\cdot\sigma^2}\,.$$

Thus, the number of samples needed to have a distinguishing advantage $1/2$ (i.e., guess correctly with probability $3/4$) is

$$n_{1/2} = \theta\frac{(4\ln 2)\cdot\sigma^2}{(x-y)^2}$$

with $0.41 \leq \theta \leq 2$.

## References


1. C. Blondeau, A. Bay, S. Vaudenay. Protecting against Multidimensional Linear and Truncated Differential Cryptanalysis by Decorrelation. In *Fast Software Encryption'15*, Istanbul, Turkey, Lecture Notes in Computer Science 9054, pp. 73–91, Springer-Verlag, 2015. Eprint 2015/380. http://eprint.iacr.org/2015/380.pdf
2. W. Dai, V.T. Hoang, S. Tessaro. Information-Theoretic Indistinguishability via the Chi-Squared Method. In *Advances in Cryptology CRYPTO'17*, Santa Barbara, California, U.S.A., Lecture Notes in Computer Science 10401–10403, pp. 497–523 vol. 3, Springer-Verlag, 2017.

3. T. van Erven, P. Harremoës. Rényi Divergence and Kullback-Leibler Divergence. *IEEE Transactions on Information Theory*, vol. IT-60 (7), pp. 3797–3820, 2014. http://arxiv.org/abs/1206.2459
4. C.A. Fuchs, J. van de Graaf. Cryptographic Distinguishability Measures for Quantum Mechanical States. *IEEE Transactions on Information Theory*, vol. IT-45 (4), pp. 1216–1227, 1999. https://arxiv.org/abs/quant-ph/9712042
5. G.L. Gilardoni. On Pinsker's and Vajda's Type Inequalities for Csiszár's $f$-Divergences. *IEEE Transactions on Information Theory*, vol. IT-55 (11), pp. 5377–5386, 2010.
6. T. Pöppelmann, L. Ducas, T. Güneysu. Enhanced Lattice-Based Signatures on Reconfigurable Hardware. In *Cryptographic Hardware and Embedded Systems CHES'14*, Busan, Korea, Lecture Notes in Computer Science 8731, pp. 353–370, Springer-Verlag, 2014.
7. T. Pöppelmann, L. Ducas, T. Güneysu. Enhanced Lattice-Based Signatures on Reconfigurable Hardware. IACR Eprint 2014/254 report, 2014. http://eprint.iacr.org/2014/254.pdf
8. R. Renner. On the Variational Distance of Independently Repeated Experiments. CoRR abs/cs/0509013, 2005. http://arxiv.org/abs/cs/0509013
9. I. Sason, S. Verdú. $f$-Divergence Inequalities. Preprint arXiv:1508.00335 [cs.IT], 2016. https://arxiv.org/abs/1508.00335
10. S. Vaudenay. Decorrelation: a Theory for Block Cipher Security. *Journal of Cryptology*, vol. 16, pp. 249–286, 2003.