A Convex Analysis Approach to Computational Entropy

Maciej Skórski

University of Warsaw

Abstract. This paper studies the notion of computational entropy. Using techniques from convex optimization, we investigate the following problems:

1. Can we derandomize the computational entropy? More precisely, for the computational entropy, what is the real difference in security defined using the three important classes of circuits: deterministic boolean, deterministic real valued, or (the most powerful) randomized ones?
2. How large the difference in the computational entropy for an unbounded versus efficient adversary can be?
3. Can we obtain useful, simpler characterizations for the computational entropy?

The first question was answered affirmatively for the most important notion of HILL entropy but was open for the metric-type computational entropy, widely used in the leakage-resilient cryptography. In this case, we show that the answer depends on what is the underlying variant of the information-theoretic entropy in the definition of the metric entropy. More precisely, the answer is negative for the commonly used min-entropy based computational entropy. Surprisingly, we show that for all other Renyi entropies the answer is positive - security given by unbounded deterministic circuits can be still much worse than that guaranteed by efficient randomized circuits.

In the second problem, we obtain some lower-bound type results. Especially, considering conditional computational entropy for two random variables \( X \in \{0, 1\}^n \) and \( Z \in \{0, 1\}^m \), we show that even if the security parameters are exponential in \( n + m \), the amount of entropy can be still noticeably higher than that seen by unbounded adversary. Also, for a fixed distribution, decreasing the security parameters by a factor \( 2^C \), can result in increasing the entropy by \( C \) bits, which agrees with intuition.

Studying the third problem, we derive a series of lemmas giving a characterization of the metric entropy for various definitions. As an example of application, we give extremely simple proofs of leakage lemmas, being a central tool in the leakage-resilient cryptography.
1 Introduction

Entropy is the fundamental concept on which information-theory is founded. Since its introduction by Shannon \cite{Sha48} the definition of entropy has been generalized in many ways, including the computational variants of this notion (introduced in \cite{Yao82} and \cite{HILL99}), which turn out to be very useful in the computational complexity theory and cryptography.

There are at least three important and different natural approaches to define computational entropy: the first one based on compressibility (“Yao entropy”), the second one based on the notion of unpredictability (“unpredictability entropy”) and the other one based on the concept of computational indistinguishability (“Metric and HILL entropy”). The relationships between these notions were studied first by Barak et al. in \cite{BSW03}; the reader might also wish to refer to \cite{Rey11} for a survey. In the recent years probably the most popular computational entropy variants were the Metric and the HILL entropies. This is partly due to the fact that this notion is often used by authors studying leakage-related problems (Dziembowski and Pietrzak \cite{DP08}, Reingold et al. \cite{RTTV08}, Reyzin and Fuller \cite{FR12}, Kai-Min Chung et al. \cite{CKLR11}, Krenn et al. \cite{KPW13}). The second important reason is that applying an extractor to a random variable having high HILL Entropy (or even Metric Entropy), one obtains a pseudorandom distribution \cite{FR12}.

1.1 Computational Entropy issues

A major difficulty with the use computational entropy is that it can be defined in many ways, depending on particular application, that often seem to be nonequivalent or not to admit a simple proof. Most often the differences come from the usage of different classes of distinguishers or because there is no standard way of defining conditional computational entropy. As a consequence, for many results in this area we do not know whether they are true if a small change in the definition is made. An example of such a situation is the notion of Metric Entropy introduced in \cite{DP08} and generalized in \cite{FR12}, reflecting in both cases the problem with determining what are the relationships between Metric Entropy computed against different classes of distinguishers: boolean deterministic, \([0,1]\)-valued or boolean randomized ones (for the HILL Entropy it is easy to show that all these classes are equivalent \cite{FR12}).

The another important issue is a very useful estimate used in the leakage-resilient cryptography, called the “leakage chain rule”, provable for restricted types of conditional computational entropy but known to be false in general \cite{KPW13}. Yet another important topic is existence of a simple characterization for the Metric Entropy in special cases. Besides of being of independent interest, such characterizations can have surprisingly powerful applications (cf. Section 7 in \cite{BSW03}). Thus, although a lot has been done, it seems that systematization of definitions and studying relations between different variants of entropy (even in most often used circuits model) is still needed. Our motivation is to contribute to this task, focusing on indistinguishability based computational entropy through this paper.

1.2 Our techniques

Our main technique is a novel and interesting observation that the concept of the computational entropy is strictly related to the separating convex sets problem. This approach turns out to be especially useful for the metric-type of computational entropy. The “extreme” distributions that satisfy the metric-entropy constraints turn out to be indeed extreme points and allow us to apply the powerful machinery of convex analysis. Especially we show that such problems as comparing the security of the metric entropy in different models of an adversary, are deeply dependent on the geometry of certain convex sets. We believe that this approach can be of independent interest.

1.3 Our results and the organization of the paper

The remainder of the paper is organized as follows. In Section 2 we give some basic notations and introduce definitions of Computational Entropy. In Section 3, by techniques similar to these used in results related to \(\delta\)-hard functions, we show a separation between Computational Entropy and Smooth Entropy (which can be viewed as comparing Computational Entropy seen by a bounded and by an unbounded adversary). In Section 4, by solving convex optimization problems, we obtain explicit characterizations of most interesting generalizations of metric entropy. As some of examples of application we reprove the classical relationship between Rényi Entropy for different orders and also give an extremely short proof of the ‘leakage lemma” and the ‘leakage chain rule’ for so called relaxed entropy. Section 5 deals with the problem of comparing Computational Metric Entropy for different classes of distinguishers used in the definition. We show that it can happen, that the deterministic unbounded adversary is much more weaker than for
the efficient randomized one. Surprisingly this is not the case of the most popular metric min-entropy. Especially, we construct a random variable $X \in \{0,1\}^n$ such that its metric collision entropy for two cases: (a) seen by deterministic unbounded adversary and (b) seen by adversary using only randomized circuits of size only $O(n)$ and accepting the distinguishing advantage to be only $1/poly(n)$, differs by $\Omega(\log \log n)$. Even more pathological result can be obtained for the Shannon Entropy: it is possible that the ‘gap’ in the amount of entropy for the unbounded deterministic and randomized adversary accepting constant distinguishing advantage, is even $\Omega(n)$.

## 2 Preliminaries

**Information-theoretic notions** The idea commonly used to define computational entropy is to generalize a convinient theoretic-infomation notion of entropy. Following this way, we start with recalling the notion of the Rényi Entropy.

**Definition 1 (Rényi Entropy).** Given a random variable $X \in \{0,1\}^n$ we say that its Rényi Entropy of order $\alpha$ (or in short: $\alpha$-Rényi Entropy) is at least $k$ if and only if

$$\|P_X\|_{\alpha-1} = \left( \mathbf{E}_{x \in X} (P_X(x))^{\alpha-1} \right)^{\frac{1}{\alpha-1}} \leq 2^{-k}$$

This definition covers also the important cases of the Collision Entropy ($\alpha = 2$), the Shannon Entropy ($\alpha \rightarrow 1$) and the Min-entropy ($\alpha \rightarrow \infty$). By calculating these limits, one can give the explicit definitions for the last two cases:

**Definition 2 (Shannon Entropy).** Given a random variable $X \in \{0,1\}^n$ we say that its Shannon entropy is at least $k$ if and only if

$$-\sum_x P_X(x) \log P_X(x) \geq k$$

where we define $p \log p = 0$ for $p = 0$.

**Definition 3 (Min-Entropy).** Given a random variable $X \in \{0,1\}^n$ we say that its Min-Entropy is at least $k$ if and only if

$$P_X(x) \leq 2^{-k} \quad \text{for all } x \in \{0,1\}^n$$

For some applications, like for the randomness extraction, the smoothed version of entropy is usefull. The key concept behind smooth entropy is that we allow $X$ to be only close (in some metric sufficiently strong to our purposes) to a distribution with required entropy, instead of expecting $X$ to satisfy the entropy constraints by itself.

**Definition 4 (Statistical Distance).** Let $X, Y \in \{0,1\}^n$ be two random variables. The statistical distance of distributions $P_X, P_Y$ is defined to be $\Delta(X,Y) = \frac{1}{2} \sum_x |P_X(x) - P_Y(x)|$.

**Definition 5 (Smooth Rényi Entropy).** Given $\varepsilon > 0$ and a random variable $X \in \{0,1\}^n$, we say that it has Smooth $\alpha$-Rényi Entropy at least $k$, if there exists a random variable $Y \in \{0,1\}^n$ such that $H_\alpha(Y) \geq k$ and $\Delta(X,Y) \leq \varepsilon$.

### 2.1 Computational Entropy

The intuition behind HILL Entropy is that we think of $X$ as having high computational entropy if it is computationally indistinguishable from a distribution with (chosen) information-theoretic entropy. The computational variant of min-entropy was introduced in [HILL99]. Below we generalize this concept replacing the min-entropy by Rényi Entropy.

**Definition 6 (Computational HILL Rényi Entropy).** Given $\varepsilon > 0$, a class of distinguishers $\mathcal{D}$ and a random variable $X \in \{0,1\}^n$, we say that $X$ has at least $k$ bits of HILL Computational Rényi Entropy of order $\alpha$ against $(\mathcal{D}, \varepsilon)$ and denote by $H_{\alpha}^{\text{HILL}, \mathcal{D}, \varepsilon}(X) \geq k$ if there exist a distribution $Y$ over $\{0,1\}^n$ satisfying $H_\alpha(Y) \geq k$ such that for any $D \in \mathcal{D}$ holds $|ED(X) - ED(Y)| \leq \varepsilon$.

Metric entropy is defined by reversing the order of quantifiers:

**Definition 7 (Computational Metric Rényi Entropy).** Given $\varepsilon > 0$ and a class of distinguishers $\mathcal{D}$ we say that the random variable $X \in \{0,1\}^n$ has at least $k$ bits of Metric Computational Rényi $\alpha$-Entropy against $(\mathcal{D}, \varepsilon)$ and denote by $H_{\alpha}^{\text{metric}, \mathcal{D}, \varepsilon}(X) \geq k$ if for any $D \in \mathcal{D}$ there exist a distribution $Y$ over $\{0,1\}^n$ satisfying $H_\alpha(Y) \geq k$ and $|ED(X) - ED(Y)| \leq \varepsilon$.

1 The distinguishers can be deterministic or randomized $[0,1]$-valued functions.
2.2 Conditional Computational Entropy

The conditional computational entropy is defined in the similar way via underlying theoretic-information entropy measure. Since there is no agreement how to define Conditional Rényi Entropy, to make this discussion clear we restrict us only to the case of min-entropy. Usually one defines the conditional min-entropy in one of the two ways:

**Definition 8 (Conditional Min Entropy).** Given a joint distribution \( (X, Z) \) we say that \( X \) conditioned on \( Z \) has min-entropy at least \( k \) and denote by \( H(X|Z) \geq k \)

\[
\forall z : H_\infty(X|Z = z) \geq k
\]

**Definition 9 (Average Conditional Min Entropy [DORS08]).** Given a joint distribution \( (X, Z) \) we say that \( X \) conditioned on \( Z \) has average min-entropy at least \( k \) and denote by \( \tilde{H}(X|Z) \geq k \) if

\[
E_{z \leftarrow Z} \left[ 2^{-H_i(X|Z = z)} \right] = E_{z \leftarrow Z} \left[ \max_x P_{X|Z,z}(x) \right] \leq 2^{-k}
\]

The conditional computational entropy is defined similarly to the unconditional case.

**Definition 10 (Conditional Computational HILL Entropy).** Given \( \epsilon > 0 \), a class of distinguishers \( D \) and a pair of random variables \( X \in \{0,1\}^n, Z \in \{0,1\}^m \) we say that \( X \) conditioned on \( Z \) has at least \( k \) bits of HILL Computational Min Entropy against \( (D, \epsilon) \) and denote by \( H_{\infty}(X|Z) \geq k \) if there exists a distribution \( (Y, Z') \) over \( \{0,1\}^{n+m} \) satisfying \( H_{\infty}(Y|Z') \geq k \) such that for any \( D \in D \) holds the inequality \( |ED(X, Z) - ED(X, Z')| \leq \epsilon \).

The conditional computational metric entropy is defined by changing the order of the quantifiers. Metric as well as HILL conditional entropy can be defined as average or non-average conditional entropy depending on use \( \tilde{H}_{\infty} \) or \( H_{\infty} \) and denoted using these symbols. For clarity we do not give the rest of possible definitions.

For the sake of completeness we note that there is a definition that allows \( Z \) to change together with \( Y \). This leads to the notion of relaxed computational min entropy:

**Definition 11 (Conditional Computational HILL Relaxed Entropy, [Rey11]).** Given \( \epsilon > 0 \), a class of distinguishers \( D \) and a pair of random variables \( X \in \{0,1\}^n, Z \in \{0,1\}^m \), we say that \( X \) conditioned on \( Z \) has at least \( k \) bits of Relaxed HILL Computational Entropy against \( (D, \epsilon) \) and denote by \( H_{HILL-rlx}(X|Z) \geq k \) if there exists a distribution \( (Y, Z') \) over \( \{0,1\}^{n+m} \) satisfying \( H_{\infty}(Y|Z') \geq k \) such that for any \( D \in D \) holds \( |ED(X, Z) - ED(X, Z')| \leq \epsilon \).

This entropy also can be considered in average or non-average aspects, with HILL or Metric type of indistinguishability and has remarkably good properties for some leakage-related problems, as we will see later.

**Relationships between HILL and Metric Entropy**

The Metric entropy, which was was introduced after the HILL one, often turns out to be more convenient in applications (for instance, to prove leakage-related results). It is known that from Metric Entropy computed against real valued (or randomized) circuits, then there exists a conversion to HILL entropy [BSW03]. This result in its full generality can be stated as follows

**Theorem 1 (Generalization of [BSW03], Thm. 5.2).** Let \( \mathcal{P} \) be the set of all probability measures over \( \Omega \). Suppose that we are given a class \( D \) of \( [0,1] \)-valued functions on \( \Omega \), with the following property: if \( D \in D \) then \( D^e \) where

\[
1 - D \in D.
\]

For \( \delta > 0 \), let \( D' \) be the class consisting of all convex combinations of length \( \lceil \log |D| / \delta^2 \rceil \) over \( D \).

Let \( \mathcal{C} \subset \mathcal{P} \) be any arbitrary convex subset of probability measures and \( X \in \mathcal{P} \) be a fixed distribution. Consider the following statements:

\begin{enumerate}[i]
    \item \( X \) is \( (D, \epsilon + \delta) \) indistinguishable from some distribution \( Y \in \mathcal{C} \) (HILL Entropy)
    \item \( X \) is \( (D', \epsilon) \) indistinguishable from the set of all distribution \( Y \in \mathcal{C} \) (Metric Entropy)
\end{enumerate}

Then (ii) implies (i).

The sketch of the proof appears in Appendix.

**Remark 1.** By choosing \( \Omega = \{0,1\}^{n+m} \), a random variable \( Z \in \{0,1\}^m \) and \( \mathcal{C} \) to be the set of all distributions \((Y, Z) \) satisfying \( (Y, Z) : H_\infty(Y|Z) \geq k \) or alternatively \( H_{\infty}(Y|Z) \geq k \), we obtain the conversion from the Metric Conditional Entropy to the HILL Conditional Entropy, for both: worst case and average case variants.
2.3 Relationship of Convex Analysis to Metric Entropy

Let us notice, that the both notions: HILL and Metric entropy can be rephrased in a geometrical language as inability to separate between two convex sets. More precisely, any distribution \( \mathbf{P}_X \) on \( \{0, 1\}^n \), after enumerating the elements of \( \{0, 1\}^n \), can be (uniquely) identified with a vector in \( \mathbb{R}^2 \). Similarly, any real valued function \( D \) on \( \{0, 1\}^n \) can be identified with a vector in the same space. Taking the expected value becomes then the scalar product

\[
\langle D, \mathbf{P}_X \rangle = \sum_x \mathbf{P}_X(x) D(x) = \mathbf{E}_{x \sim \mathbf{P}_X} D(x)
\]

By considering the min-entropy, for instance, it is easy so see that \( H_{\infty}^{\text{Metric-D,} \epsilon}(X) < k \) if and only if there exists \( D \in \mathcal{D} \) such that for all \( \mathbf{P}_Y \in \mathcal{Y} \)

\[
|\langle D, \mathbf{P}_X - \mathbf{P}_Y \rangle| = \left| \sum_x D(x) (\mathbf{P}_X(x) - \mathbf{P}_Y(x)) \right| \geq |ED(X) - ED(Y)| \geq \epsilon
\]

where \( \mathcal{Y} \) is the set of all distributions on \( \{0, 1\}^n \)s with min-entropy at least \( k \). We will see later that the absolute value above can be removed by considering classes \( \mathcal{D} \) which are closed under complements (i.e. if \( D \in \mathcal{D} \) then also \( 1 - D \in \mathcal{D} \)). Then we get the inequality \( \langle D, \mathbf{P}_X \rangle \geq \langle D, \mathbf{P}_Y \rangle + \epsilon \) valid for all \( \mathbf{P}_Y \in \mathcal{Y} \). Thus defining the metric entropy is nothing more than just saying that a given distribution \( X \) cannot be separated (in the sense known from functional analysis or convex analysis) from the set \( \mathcal{Y} \) (i.e. from all its elements at once). In the other hand, \( D \) that contradicts the definition is exactly a separating hyperplane. Hence, methods of convex analysis can be applied to study the properties of metric-type entropies. The HILL-type definition is less compatible with this approach, as it is a bit stronger assumption, namely that we are not able to separate any pair \( \mathbf{P}_X, \mathbf{P}_Y \) where \( \mathbf{P}_Y \in \mathcal{Y} \). In this paper we follow the terminology used in computer science, saying about distinguishing instead of separating as in math.

2.4 Used conventions and important remarks

Through this paper we will use mostly the already defined computational min-entropy, saying in short about computational entropy based computational entropy. We also use the following natural convention: replacing \( \mathcal{D} \) by a pair \((\{0, 1\}, s)\) or \(((0, 1), s)\) if we mean deterministic circuits of size \( s \) respectively boolean and \([0, 1]\)-valued. Writing \((\text{rand}(0, 1), s)\) in the place of \( \mathcal{D} \) we mean randomized boolean circuits of size \( s \). If the circuit size \( s \) is omitted in the description of a circuit class, it is assumed to be unbounded. For the boolean function \( D \) we denote \( |D| = \sum_{x \in \text{dom} D} D(x) \).

Note that although one can define and use computational entropy based on the Rényi Entropy of any arbitrary order, using of min-entropy as a reasonable compromise between the the convenience of analysis and preserving so much generality as possible, is not a big restriction in practice, as long as one uses real valued distinguishers. To pass between Renyi Entropies for different order, one uses the fact that the values of the Smooth Rényi Entropy for different order cannot be differ more than a small additive constant. The precise statement is given bellow:

Lemma 1 ([RW]). Suppose that \( X \) is a distribution over \( \{0, 1\}^n \). Then for \( \alpha > 1 \)

\[
H_\alpha^{\infty}(X) \geq H_\alpha(X) - \frac{1}{\alpha - 1} \log \frac{1}{\epsilon}
\]

We will obtain another proof of this result using a characterization of computational metric entropy. These equivalence does not cover the Shannon Entropy case. It is worth of noting that the Shannon Entropy based Computational Entropy also found applications where it becomes more suitable than the computational min-entropy [VZ12].

3 Separation between Computational Entropy and Smooth Entropy

In this section we examine the existence of a conversion rule from the computational to the smooth entropy:

Suppose that \( H_{\infty}^{\text{Metric-D,} s, \epsilon}(X|Z) \geq k \), where \( X \in \{0, 1\}^n \), \( Z \in \{0, 1\}^m \). What are the conditions on \( s, \epsilon \) that guarantee that \( H_{\infty}^{\epsilon'}(X|Z) \geq k' \) with \( \epsilon' \leq 2C \epsilon \) and \( k' \geq k - C \) for some constant \( C \)?
In Section 5 we will prove that if the security parameters are sufficiently strong, more precisely, if \( s = \mathcal{O}(2^{k+m}) \), then the computational min-entropy becomes the smooth entropy. For the unconditional case also exponentially small \( \epsilon \) is sufficient (see Section 5 Corollary 4). Interestingly, this result can be inverted. In this section we show that

Given the metric entropy of \( X \mid Z \) one really needs the security to be exponentially strong in \( k + m \), to obtain the smooth entropy with comparable parameters.

We stress that although the existence of a separation between the Metric and Smooth Entropy is almost obvious, the quantitative bound which is exponential in both: \( k \) and \( m \) is less trivial to see. Since the maximal entropy of \( X \mid Z \) is \( n \), it follows that even if distinguishers were given access to an oracle over \( \{0,1\}^n \), the entropy could be still non-trivial.

**Remark 2.** Since Min-Entropy is the smallest one among other Renyi Entropies, and because there is efficient conversion for Smooth Renyi Entropies (Lemma 1) it is sufficient to consider the case of Min Entropy.

**Separation for unconditional Computational Min-entropy**

**Theorem 2.** For any \( C > 1 \), there exists \( X \) such that \( H^\text{Metric, O}(2^k, 2^{\log(2^k)})/C \) \( H^\text{C} \) \( k + C \) but \( H^\text{C} \) \( k + 1 \).

**Proof.** The main idea is to reduce the problem to a problem of approximating of a certain function, which will turn out to be hard for limited size circuits. Let \( A \) and \( S \) be sets of cardinality \( 2^k \) and \( 2^{k+C} \) and \( A \subset S \). Denote \( B = S \setminus A \).

Consider the random variable \( X = U_A \). It is easy to see that \( H^\text{Metric}_\infty \) \( k + 1 \). Observe that

\[
ED(X) - ED(U_S) = ED(U_A) - ED(U_S) = (1 - 2^{-C}) \left( \mathbb{P}(D(U_A) = 1) - \mathbb{P}(D(U_B) = 1) \right) = (1 - 2^{-C}) \left( \mathbb{P}(D(U_A) = 1) + \mathbb{P}(D(U_B) = 0) \right) - (1 - 2^{-C}),
\]

hence, assuming \( H^\text{Metric, det}_{\{0,1\}, s, C} \) \( k + C \) for \( \epsilon = \delta(1 - 2^{-C}) \), we get

\[
\mathbb{P}(D(U_A) = 1) + \mathbb{P}(D(U_B) = 0) > 1 + \delta
\]

The proof easily follows now from the following lemma, being a strenghtening of the classical result on the existence of \( \delta \)-hard functions.

**Lemma 2.** For any \( C \geq 1 \) and sufficiently large \( \ell \) there exists a boolean function \( f \) over \( \{0,1\}^\ell \), such that \( \text{bias}(f) = 1 - 2^{1-C} \) and for all circuits \( D \) of size \( \mathcal{O}(2^{\ell-C}S^2/\ell-C-2\log(1/\delta)) \) we have

\[
\mathbb{P}_{x \leftarrow f^{-1}(\{1\})}(D(x) = f(x)) + \mathbb{P}_{x \leftarrow f^{-1}(\{0\})}(D(x) = f(x)) < 1 + \delta
\]

The proof follows by a standard application of the Chernoff Bound and the union bound over all circuits of bounded size. See Appendix 2 for the details and discussion.

**Separation for Conditional Computational Min Entropy**

**Theorem 3.** For sufficiently large \( n \), and for any \( C > 0 \), \( k < n - C \) and \( \epsilon > 0 \) there exists a pair of jointly distributed random variables \( X \in \{0,1\}^n \), \( Z \in \{0,1\}^m \) such that

(i) \( H_{\infty}^1(X \mid Z) \leq k + 1 \)
(ii) \( H^\text{Metric, det}_{\{0,1\}, s, C}(X \mid Z) \geq k + C \) for \( s = \mathcal{O}\left(\frac{2^{k+m}}{(k+m)\log(2^{k+m}/\epsilon)}\right) \)

The proof is longer than for the unconditional case. The key point is that in the conditional case it is significantly harder to find an appropriate "hard approximation task". See the proof of Theorem 3 in the Appendix.
4 Characterizations of Metric Entropy

**General Characterization Theorem** The following result can be viewed as a general characterization of Metric Entropy. The easy proof is given in the Appendix.

**Theorem 4.** Let \( \mathcal{D} \) be a class of real valued functions on \( \{0,1\}^n \) closed under complements, \( \mathcal{Y} \) be a non-empty compact convex set of probability distributions over \( \{0,1\}^n \) and \( X \in \{0,1\}^n \) be a random variable. Then the following conditions are equivalent:

(i) For every \( Y \in \mathcal{Y} \) there exists \( D \in \mathcal{D} \) such that \( |\mathcal{E}D(X) - \mathcal{E}D(Y)| < \epsilon \)

(ii) For every \( D \in \mathcal{D} \) we have \( \mathcal{E}D(X) \leq \max_{Y \in \mathcal{Y}} \mathcal{E}D(Y) + \epsilon \)

**Remark 3.** If the entropy is defined based on underlying information-theoretic entropy measure \( H \), for instance Renyi Computational Entropy defined in Section 2, then the set \( \mathcal{Y} \) is just so called superlevels set: it consists of all distributions having the (information-theoretic) entropy at least \( k \). For the conditional relaxed entropy of \( X|Z \), we set \( \mathcal{Y} = \{ P_{Y,Z} : Y \in \{0,1\}^n, Z' \in \{0,1\}^m, H_\infty(X|Z') \geq k \} \).

It is clear that we need to solve the maximization task explicitly, in order to obtain a characterization for a concrete variant of Metric Entropy.

**Renyi Entropy**

By computing \( \max_{Y \in \mathcal{Y}} \mathcal{E}D(Y) \) in Theorem 4 we characterize the most important cases of the Metric Rényi Entropy.

**Lemma 3.** Let \( \alpha > 1 \) be fixed, let \( D : \{0,1\}^n \to \{0,1\} \) be a function and \( \mathcal{Y}_k = \{ Y \in \{0,1\}^n : H_\alpha(Y) \geq k \} \). Then

\[
\max_{Y \in \mathcal{Y}_k} \mathcal{E}D(Y) = \begin{cases} 
 p_D \cdot |D|, & \text{if } |D| < 2^k \\
 1, & \text{otherwise}
\end{cases}
\]

(1)

where \( p_D \), for \( |D| \leq 2^k \), is the greatest number satisfying the following system

\[
\begin{align*}
 p_D^\alpha |D| + q_D^\alpha |D^c| &= 2^{-(\alpha - 1)k} \\
p_D |D| + q_D |D^c| &= 1 \\
p_D, q_D &\geq 0
\end{align*}
\]

(2)

Moreover, the solution \( p_D \) is unique provided that \( k < n - 1 \).

The proof is not hard but technical and is left to the appendix (see the proof of Theorem 4). Especially, for the cases \( \alpha \to 1 \), \( \alpha = 2 \) and \( \alpha \to \infty \) corresponding to the Shannon, Collision and Min-Entropy respectively, after some calculations we obtain the following characterizations:

**Corollary 1** (Metric Shannon Entropy). Let \( \mathcal{D} \) be a class of boolean functions closed under complements. Then the following conditions are equivalent

(i) \( H_1^{\text{Metric},D,\epsilon}(X) \geq k \)

(ii) For every \( D \in \mathcal{D} \) such that \( |D| \leq 2^k \), \( \mathcal{E}D(X) \leq p_D |D| + \epsilon \) holds for \( p_D \) solving the system

\[
\begin{align*}
 -p_D |D| \log p_D - q_D |D^c| \log q_D &= k \\
p_D |D| + q_D |D^c| &= 1 \\
p_D, q_D &\geq 0
\end{align*}
\]

(3)

**Corollary 2** (Metric Collision Entropy). Let \( \mathcal{D} \) be a class of boolean functions closed under complements. Then the following conditions are equivalent

(i) \( H_2^{\text{Metric},D,\epsilon}(X) \geq k \)

(ii) The inequality \( \mathcal{E}D(X) \leq p_D |D| + \epsilon \) holds for every \( D \in \mathcal{D} \), and \( p_D \) given by

\[
p_D = 2^{-n} + \sqrt{|D^c| |D|^{-1} (2^{-k-n} - 2^{-2n})}
\]

(4)
Corollary 3 (Metric Min-Entropy, [BSW03]). Let $\mathcal{D}$ be a class of boolean functions closed under complements. Then the following conditions are equivalent

(i) $H_{\text{Metric}, \mathcal{D}, \epsilon}(X) \geq k$
(ii) The inequality $\mathbb{E} D(X) \leq 2^{-k|D|} + \epsilon$ holds for every $D \in \mathcal{D}$

Note that from the characterization in Lemma 3 and Lemma it follows that one need to check only distinguishers of size $\exp(k)$ to prove that the metric entropy is $k$.

Corollary 4. Let $X \in \{0, 1\}^n$ be a random variable, $\alpha \in [1, \infty]$ and $s = \Omega(2^k \alpha)$. Then $H_{\alpha}^{\text{Metric}, \{0, 1\}, s} = H_{\alpha}^{\text{Metric}, \{0, 1\}, \infty}$.

Relaxed Computational Entropy

Lemma 4. Let $X \in \{0, 1\}^n$, $Z \in \{0, 1\}^m$ be random variables, and let

$$\mathcal{Y} = \{P_{Y, Z'} : Y \in \{0, 1\}^n, Z' \in \{0, 1\}^m, H_\infty(Y|Z') \geq k\}$$

Then for every boolean function $D$ on $\{0, 1\}^{n+m}$ we have

$$\max_{P_{Y, Z'} \in \mathcal{Y}} \mathbb{E} D(Y, Z') = 2^{-k} \min \left( \max_z |D(\cdot, z)|, 1 \right)$$

Proof. Observe that for every $P_{Y, Z'} \in \mathcal{Y}$ we have

$$\mathbb{E} D(Y, Z') \leq \sum_{x, z} P_{X, Z}(x, z) D(x, z)$$

$$\leq \sum_{x, z} 2^{-k} P_{Z}(z) D(x, z)$$

$$\leq 2^{-k} \max_z |D(\cdot, z)|$$

Let $z'$ be chosen over $z \in \{0, 1\}^m$ so that it maximizes $|D(\cdot, z)|$. The equality in the estimate above is achieved provided that $|D(\cdot, z')| \leq 2^k$, $Z'$ is a point mass distribution at $z'$ and $Y|Z' = z'$ satisfies $P_{Y|Z' = z'} = 2^{-k}$ if $D(x, z) = 1$. For the case $\max_z |D(\cdot, z)| > 2^k$, let $Y'|z' = z'$ be a uniform distribution on arbitrary $2^k$-element subset of $\{x : D(x, z) = 1\}$. Then we have $\mathbb{E} D(Y, Z') = 1$.

Corollary 5 (Relaxed Metric Entropy). Let $\mathcal{D}$ be a class of boolean functions closed under complements. Then the following are equivalent

(i) $H_{\text{Metric-rlx}, \mathcal{D}, \epsilon}(X|Z) \geq k$
(ii) The inequality $\mathbb{E} D(X, Z) \leq 2^{-k} \max_z |D(\cdot, z)| + \epsilon$ holds for every $D \in \mathcal{D}$

4.1 Examples of Applications

As a first example we give below much simpler proofs of the leakage chain rule for relaxed-type entropy and for the leakage lemma. Interestingly, there is no hope for proving an efficient version (meaning a bound on loss in security parameters) for non-relaxed definition, as shown recently by Krenn et al. [KPW13].

Theorem 5 (Leakage Lemmas). Let $X, Z_1, Z_2$ be correlated random variables taking values in $\{0, 1\}^n$, $\{0, 1\}^m$ and $\{0, 1\}^s$ respectively. Then the following estimate, called “the chain rule”, is true [GW10, Rey11]

$$H_{\infty}^{\text{Metric-rlx}, \{0, 1\}, s, 2^{m+2}\epsilon} (X|Z_1, Z_2) \geq H_{\infty}^{\text{Metric-rlx}, \{0, 1\}, s, \epsilon} (X|Z_1) - m_2.$$

Especially, for the uncondition case $Z_1 = \emptyset$ we obtain the so-called “leakage lemma” [DP08, RTTV08, FR12]

$$H_{\infty}^{\text{Metric}, \{0, 1\}, s, 2^{m+2}\epsilon} (X|Z_2) \geq H_{\infty}^{\text{Metric}, \{0, 1\}, s, \epsilon} (X) - m_2.$$
Theorem 1 giving conversion between Metric and HILL entropy we know that real-valued distinguishers are equivalent.

Corollary 6. In the second example, we reprove a result on passing between Renyi Entropy for different orders.

Remark 4. Note that both results are often formulated using the HILL entropy, with the weaker, by a factor poly (1/ε), security parameter s. This factor is exactly the cost of the conversion from Metric to HILL entropy (Theorem 1) can be applied for the relaxed-metric entropy, because for this notion boolean and real valued circuits are equivalent as will see later; the same is true for the conditional metric worst-case entropy which is known to be equivalent to the conditional metric average-case entropy up to loss log(1/ε) in the entropy ammount). Sometimes loss appears in ε instead in s, which can be also thought as an equivalent statement. Thus, our proof really implies the original results.

In the second example, we reprove a result on passing between Renyi Entropy for different orders.

Corollary 6. The proof of Lemma 1

Proof. Since p is a solution of the equation \( p^\alpha |D| + \left( \frac{1-p|D|}{|D|} \right)^\alpha = 2^{-(\alpha - 1)k} \), we have \( p|D| < \left( \frac{|D|}{2^k} \right)^{\frac{\alpha - 1}{\alpha}} \). Suppose that \(|D| > 2^k \cdot 2^{-C\alpha}\) where the parameter C will be specified later. Then \( \left( \frac{|D|}{2^k} \right)^{\frac{\alpha - 1}{\alpha}} < \frac{|D|}{2^k} \). In turn if the opposite inequality holds then \( \left( \frac{|D|}{2^k} \right)^{\frac{\alpha - 1}{\alpha}} \leq 2^{-C(\alpha - 1)} \). In any case, we get \( p|D| \leq |D| + \epsilon' \) where \( k' = k - C \) and \( \epsilon' = 2^{-C(\alpha - 1)} \). Thus the distribution X has a metric entropy at least \( k' \) with error \( \epsilon' \) against all boolean functions. According to Theorem 1 in the next section this is the same entropy as if \([0,1]\) valued functions would be used. From Theorem 1 giving conversion between Metric and HILL entropy we know that X has HILL Entropy with the same parameters. Finally, HILL Entropy against all \([0,1]\) valued function is clearly the same as Smooth Entropy. Choosing \( C = \frac{1}{\alpha - 1} \log \frac{1}{\epsilon} \) we recover the estimate on the Smooth Rényi Entropy for different orders given in Lemma 1.

5 Metric Entropy Against Deterministic and Randomized Adversary

It is well known that for the HILL-type entropy there is no matter whether we use deterministic or randomized (or real valued) class of distinguishers. The reason is that we can just fix an ‘optimal’ choice of coins for a randomized function distinguishing between two probability distributions. However, this argument fails in the case of a metric-type definition because of a different order of quantifiers in the definition. So the following problem appears:

Let \( H^{\text{Metric}} \) be metric-type computational entropy (for instance, based on the Renyi Entropy of fixed order \( \alpha \)). Suppose that \( H^{\text{Metric}, \det, \{0,1\}, \alpha, \epsilon'}(X) \geq k \). Can we obtain a good lower bound on \( H^{\text{Metric}, \det, \{0,1\}, \alpha', \epsilon'}(X) \) or on \( H^{\text{Metric}, \rand, \{0,1\}, \alpha', \epsilon'}(X) \) ?

5.1 Positive answer for min-entropy

We show that even in the conditional case, for the min-entropy based metric and relaxed-metric entropy, boolean and real-valued distinguishers are equivalent.
Theorem 6. Let $X \in \{0,1\}^n$ and $Z \in \{0,1\}^n$ be random variables. Then we have $H_{\infty}^{\text{Metric, det}, \{0,1\}, s, \epsilon}(X|Z) = H_{\infty}^{\text{Metric, det}, \{0,1\}, s', \epsilon}(X|Z)$ where $s' \approx s$.

Theorem 7. Let $X \in \{0,1\}^n$ and $Z \in \{0,1\}^n$ be random variables. Then we have $H_{\infty}^{\text{Metric-rlx}, \{0,1\}, s, \epsilon}(X|Z) = H_{\infty}^{\text{Metric-rlx}, \{0,1\}, s', \epsilon}(X|Z)$ where $s' \approx s$.

The idea of the proof is to rephrase the problem as a task of separating convex sets, as discussed in Section 2.3. The standard proof by reduction requires to construct a boolean distinguisher from the given possibly real-valued one. In terms of convex analysis, it becomes a task of finding an appropriate (satisfying some restrictions) separating hyperplane. Technically this is done by calculating Lagrange Multipliers. The details are given in the Appendix, Theorem 5 and 6. Passing further to randomized circuits can be realized (with a loss) using Theorem 1.

5.2 Negative results for Renyi Entropy of order $\alpha < \infty$

Having shown some positive results for the min-entropy based metric entropy, we will show a surprising property: for any other Rényi Entropy, there exists a random variable such that its entropy against deterministic boolean circuits is strictly smaller that the entropy against real-valued circuits (and therefore also randomized circuits). Before we show the actual proof, let us give some geometric intuitions why is the min-entropy so special. The reason is, that the set of all distributions having min-entropy at least $k$, after encoding probabilities as vectors, is given by linear inequalities of the form $0 \leq p_i \leq 2^{-k}$ and $\sum_i p_i = 1$. Since that all inequality constraints form a hypercube whose faces are given by $0 - 1$ vectors, they “match” perfectly to the boolean distinguishers very well. Compare this to the collision entropy, where the entropy (collision) constraint is $\sum_i p_i^2 \leq 2^{-k}$ and the corresponding shape is clearly an ellipsoid.

From the characterization given in Lemma 4 we immediately obtain

**Proposition 1.** Let $\mathcal{Y}_k$ be the set of all distributions over $\{0,1\}^n$ with the $\alpha$-entropy at least $k$. Then the set of distributions $X$ over $\{0,1\}^n$ which are $\epsilon$-nonindistinguishable from $\mathcal{Y}_k$ by boolean functions (i.e. $H_{\text{Metric, det}, \{0,1\}, \epsilon}(X) \geq k$) is described by the following system of inequalities

$$P_X : \langle D, P_Y \rangle \leq p_D|D| + \epsilon \quad D \in \mathcal{D}, \ p_D \text{ satisfies } (2)$$

**Corollary 7.** For $1 \leq \alpha < \infty$ there exist a random variable $X \in \{0,1\}^n$ and $\epsilon_n > 0$ such that

$$H_{\alpha}^{\text{Metric, det}, \{0,1\}, 0}(X) > H_{\alpha}^{\text{Metric, rand}, \{0,1\}, \epsilon_n}(X)$$

**Proof.** Note that since the class $\det \{0,1\}$ of all boolean functions on $\{0,1\}^n$ is finite and since for every $D$ there is only finitely many solutions $p_D$, the set of solutions of (2) is a convex polyhedron on the simplex of all probability measures over $\{0,1\}^n$ thought as a subset of the space $\mathbb{R}^n$. On the other hand, the set $\mathcal{Y}$ of all distributions $Y$ with entropy $H_{\alpha}(Y)$ at least $k$, being its subset, cannot be a polyhedron as it is defined by the smooth function $(v \to \sum_i v_i^\alpha$ if $1 < \alpha < \infty$ and $v \to \sum_i v_i \log v_i$ for $\alpha = 1$). Therefore it must be strictly smaller. Thus there is a distribution $X \notin \mathcal{Y}$ such that $H_{\alpha}^{\text{Metric, det}, \{0,1\}, 0}(X) \geq k$ for some small number $\epsilon_n$. Since $P_X \notin \mathcal{Y}$, it can be strictly separated from $\mathcal{Y}$, i.e. for some $[0,1]$-valued function $D$ we have $ED(X) - ED(Y) \geq \epsilon_n$ for all $Y \in \mathcal{Y}$. But this function can be simulated by a randomized boolean circuits with arbitrary small absolute error (coming from a finite precision of the computation), let us say with error at most $\frac{1}{2}\epsilon_n$. It remains to observe, that according to the definition it means $H_{\alpha}^{\text{Metric, rand}, \{0,1\}, \epsilon_n/2}(X) < k$.

This result does not show how large the gap for the metric entropy, being seen by a deterministic or randomized adversary, can be. It is not even clear if there is a difference between a deterministic unbounded circuits and efficient but randomized ones. We provide concrete separation results for the two cases: the Shannon and the collision entropy.
Proof. We have Claim. Hence, Proof.

**Theorem 8.** For every \( k \leq n - 2 \), there exists a random variable \( X \in \{0, 1\}^n \) such that \( H^\text{Metric, \det, \{0, 1\}, 0}_2(X) \geq k \) but \( H_2(X) \leq k - \Omega(\log k) \). Moreover, we have \( H^\text{Metric, \rand, \{0, 1\}, n + \poly(k), \Theta(2^{-t})}_2(X) \leq k - \Omega(\log k) \).

**Remark 5.** The proof gives us actually the separation even between deterministic and real-valued circuits.

**Proof.** Fix a number \( k \leq n - 2 \). For \( d = 1, \ldots, 2^k \) let \( D \) be a boolean function such that \( |D| = d \) and \( p(d) = p_D \) where \( p_D \) is given by \( \square \). The sequence \( p(d) \) is well defined as the solutions \( p_D \) of \( \square \). Let \( \gamma(d) = p(d) \cdot d \). Then

\[
\gamma(d) = 2^{-n}d + \sqrt{(2^n - d) d(2^{-k-n} - 2^{-2n})},
\]

(16)

Consider the set \( S = \{ x \in \{0, 1\}^n : x = (w, 0^{n-k}) \text{ for some } w \in \{0, 1\}^k \} \) (an injection of \( \{0, 1\}^k \) into \( \{0, 1\}^n \)). Enumerate its elements by \( x^1, x^2, \ldots \) where \( x^d \) starts with the \( k \)-digit binary expansion of \( d - 1 \) and define

\[
P_X(x^1) = \gamma_1, \quad P_X(x^d) = \gamma_d - \gamma(d - 1) \quad \text{for } d = 2, \ldots, 2^k, \quad P_X(x) = 0 \text{ if } x \notin S
\]

(17)

Extend \( \gamma(d) \) by the same formula to \( d \in [1, 2^k] \). We will make use of the following properties of \( \gamma \)

**Claim.** The function \( \gamma(d) \), extended to \( d \in [1, 2^k] \), is increasing and concave.

**Proof.** We have \( \frac{\partial \gamma}{\partial d} = 2^{-n} + \frac{2^n - 2d}{2} \cdot (d(2^n - d))^{-1/2} \cdot A \) and \( \frac{\partial^2 \gamma}{\partial d^2} = -4^{-n-1} (d(2^n - d))^{-3/2} \cdot A \), where \( A = (2^{-k-n} - 2^{-2n})^{1/2} \). Thus, \( \gamma \) is increasing if \( d \leq 2^n - 1 \) and concave for \( d \leq 2^n \).

Since \( \gamma(d) \) decreases with \( d \) and \( \gamma(2^k) = 1 \), it follows that \( P_X \) is a probability measure on \( \{0, 1\}^n \). Next, we calculate the metric colision entropy and the colision entropy of \( X \).

**Claim.** We have \( H^\text{Metric, \det, \{0, 1\}, 0}_2(X) \geq k \).

**Proof.** Since \( \gamma(d) \) is a concave function, the sequence \( \gamma(d) - \gamma(d - 1) = P_X(x^d) \) is decreasing. Using this, for any boolean function \( D \) we obtain

\[
\mathbb{E}D(X) = \sum_x P_X(x) \cdot D(x)
\]

\[
\leq \max_{I \subseteq \{0, 1\}^n : |I| = |D|} \sum_{i \in I} P_X(x^i)
\]

(19)

\[
\leq \sum_{i=1}^d P_X(x^i)
\]

(20)

\[
= \gamma_d = p(d) \cdot |D|
\]

(21)

and by the characterization in Corollary \( \square \) the claim follows.

**Claim.** We have \( H_2(X) \leq k - \Omega(\log k) \)

**Proof.** Observe that

\[
P_X(x^d) = 2^{-n} + \left( \sqrt{d(2^n - d)} - \sqrt{(d-1)(2^n-d+1)} \right) \sqrt{2^{-k-n} - 2^{-2n}}.
\]

(22)

Hence,

\[
\sum_x P_X(x)^2 = 2 \cdot 2^{-n} - 2^{k-2n} + (2^{-k-n} - 2^{-2n}) \sum_{d=1}^{2^k} \left( \sqrt{d(2^n - d)} - \sqrt{(d-1)(2^n-d+1)} \right)^2.
\]

(23)
Note that
\[
\sqrt{d(2^n - d)} - \sqrt{(d-1)(2^n-d+1)} = \frac{2^n - 2d + 1}{\sqrt{d(2^n - d)} + \sqrt{(d-1)(2^n-d+1)}}
\] (24)
\[
= \Theta\left(d^{-1/2}(2^n-d)^{-1/2}(2^n-2d)\right)
\] (25)

Using this we obtain
\[
\sum_x P_X(x)^2 = 2 \cdot 2^{-n} - 2^{k-2n} + (2^{-k-n} - 2^{-2n}) \Theta\left(\sum_{d=1}^{2^k} (2^n-d)^{-1} d^{-1} (2^n-d)^2\right)
\] (26)
\[
= \Theta\left(2^{-n}\right) + \Theta\left(2^{-k-n}\right) \Theta\left(\sum_{d=1}^{2^k} \left(d^{-1} (2^n-d) + (2^n-d)^{-1} d - 2\right)\right)
\] (27)
\[
= \Theta\left(2^{-n}\right) + \Theta\left(2^{-k-n}\right) \Theta\left(2^{-k}\right).
\] (28)

Since \(\Theta\left(2^{-k}\right) = 2^{-k+\Theta(\log k)+\Theta(1)}\), the result follows.

By combining the last two claims we obtain the first part of the theorem.

**Claim.** We have \(H_2^{\text{Met, rand}\{0,1\}, n + \text{poly}(k), \Theta(2^{-k})}(X) \leq k - \Omega(\log k)\)

**Proof.** Let \(D\) be a real valued (!) function defined as \(D(x) = P_X(x)\). Let \(c\) be a positive constant (to be determined later). For every distribution \(Y\) over \([0,1]^n\) satisfying \(H_2(Y) \geq k - c \log k\), by applying the Cauchy-Schwarz Inequality and using the estimate on the collision entropy of \(X\), we obtain
\[
ED(X) - ED(Y) = \sum_x P_X(x) \cdot D(x) - \sum_x P_Y(x) \cdot D(x)
\] (29)
\[
\geq \sum_x P_X(x) \cdot D(x) - \left(\sum_x P_Y(x)^2\right)^{1/2} \left(\sum_x D(x)^2\right)^{1/2}
\] (30)
\[
= \sum_x P_X(x)^2 - \left(\sum_x P_Y(x)^2\right)^{1/2} \left(\sum_x P_X(x)^2\right)^{1/2}
\] (31)
\[
= 2^{-H_2(X)} \left(1 - 2^{H_2(Y)/2 - H_2(Y)/2}\right)
\] (32)
\[
= \Theta\left(2^{-k}\right) \left(1 - 2^{-\Theta(\log k) + c \log k}\right)
\] (33)

which is \(\Theta\left(2^{-k}\right)\) provided that \(c\) is sufficiently small. We will show how to simulate \(D\) with a randomized efficient boolean circuit \(D'\). Let \(\ell\) be chosen so that \(2^{-\ell} \ll 2^{-k}\), for instance \(\ell = \Omega(k)\).

It is easily seen that for \(x = x^d\) we have \(E_{b_1, \ldots, b_{\ell-\{0,1\}}^\ell} D'(x) = \sum_{j=1}^{\ell} 2^{-j} r_j\). Therefore, for every \(x \in \{0,1\}^n\) we have \(|E_{b_1, \ldots, b_{\ell-\{0,1\}}^\ell} D'(x) - D(x)| \leq 2^{-\ell-1}\). Since \(b_j\) are independent from \(X\) and \(Y\) it follows that
\[
ED'(X) - ED'(Y) \geq ED'(X) - ED'(Y) - 2^{-\ell-1}.
\] (34)

Hence, for all \(Y\) over \([0,1]^n\) with \(H_2(Y) \geq k\) we have
\[
ED'(X) - ED'(Y) = \Theta\left(2^{-k}\right).
\] (35)

Finally, note that the complexity of \(D'\) is at most \(O(n + \text{poly}(\ell)) = O(n + \text{poly}(k))\).

**Corollary 8.** There exists a random variable \(X \in \{0,1\}^n\) such that:

(i) \(X\) has the collision metric entropy \(k = \Theta(\log n)\) against all deterministic boolean functions, with \(c = 0\)

(ii) \(X\) has the collision metric entropy \(k - \Omega(\log \log n)\) against randomized circuits of size \(O(n)\), with \(c = \text{poly}(1/n)\)
Algorithm 1 Distinguishing between $X$ and distributions $Y$ with $H_2(Y) \geq k - \Theta(\log k)$

Require: $x \in \{0, 1\}^n$
Ensure: $D'(x)$

1: if $x \not\in S$ then
2:  return 0
3: else
4:  \hspace{1em} $d \leftarrow$ a number such that $x = x^d$
5:  \hspace{1em} for $j \leftarrow 1, \ldots, \ell$ do
6:  \hspace{2em} $r_j :=$ the $j$-th digit of the binary expansion of $P_X(x^d)$
7:  \hspace{2em} $b_j \leftarrow \{0, 1\}$ at random (flip a coin)
8:  \hspace{1em} end for
9:  \hspace{1em} $j \leftarrow$ the smallest number such that $b_j = 1$ or 0 if does not exist
10: return $r_j$
11: end if

Shannon entropy- even larger gap

For the Metric Shannon Entropy we provide the following even more stronger separation between randomized and deterministic distinguishers for the Shannon Entropy:

Corollary 9. For some absolute constant $c \in (0, 1)$, for every $n$ there exists a random variable $X \in \{0, 1\}^n$ such that:

(i) Metric Shannon Entropy of $X$ is $k \geq cn$, against all deterministic boolean functions and $\epsilon = 0$
(ii) Metric Shannon Entropy of $X$ is $k - \Omega(n)$, against all randomized circuits and $\epsilon = \Omega(1)$.

The proof is long and requires a lot of technical calculations, thus is left to the Appendix.

6 Conclusions

We developed a new “geometric” way of looking at metric-type computational entropy and show that it can be useful in some important situations, especially for the leakage-resilient cryptography. Although the tools of convex analysis seems to be complicated and unintuitive, they can yield some powerful results as we demonstrated having solved the problem of the derandomization of generalized metric-type entropy. We believe that this nonstandard approach can be helpful in improving our understanding of the computational entropy.

7 Acknowledgements

I would like to express special thanks to Stefan Dziembowski and Krzysztof Pietrzak, for their helpful suggestions and discussions.

References

BSW03. Boaz Barak, Ronen Shaltiel, and Avi Wigderson, Computational analogues of entropy., RANDOM-APPROX (Sanjeev Arora, Klaus Jansen, José D. P. Rolim, and Amit Sahai, eds.), Lecture Notes in Computer Science, vol. 2764, Springer, 2003, pp. 200–215.
CKLR11. Kai-Min Chung, Yael Tauman Kalai, Feng-Hao Liu, and Ran Raz, Memory delegation, Cryptology ePrint Archive, Report 2011/273, 2011, http://eprint.iacr.org/.
DORS08. Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith, Fuzzy extractors: How to generate strong keys from biometrics and other noisy data, SIAM J. Comput. 38 (2008), no. 1, 97–139.
DP08. Stefan Dziembowski and Krzysztof Pietrzak, Leakage-resilient cryptography in the standard model, IACR Cryptology ePrint Archive 2008 (2008), 240.
FR12. Benjamin Fuller and Leonid Reyzin, Computational entropy and information leakage, Cryptology ePrint Archive, Report 2012/466, 2012, http://eprint.iacr.org/.
GW10. Craig Gentry and Daniel Wichs, Separating succinct non-interactive arguments from all falsifiable assumptions, Cryptology ePrint Archive, Report 2010/610, 2010, http://eprint.iacr.org/.
A Proofs

Theorem 1 Let \( P \) be the set of all probability measures over \( \Omega \). Suppose that we are given a class \( D \) of \([0, 1]\)-valued functions on \( \Omega \), with the following property: if \( D \in D \) then \( D^c = \{1 - D \in D\} \). For \( \delta > 0 \), let \( D' \) be the class consisting of all convex combinations of length \( O\left(\frac{\log |\Omega|}{\delta}\right) \) over \( D \). Let \( C \subset P \) be any arbitrary convex subset of probability measures and \( X \in P \) be a fixed distribution. Consider the following statements:

i. \( X \) is \((D, \epsilon + \delta)\) indistinguishable from some distribution \( Y \in C \) (HILL Entropy)

ii. \( X \) is \((D', \epsilon)\) indistinguishable from the set of all distribution \( Y \in C \) (Metric Entropy)

Then (ii) implies (i).

Proof. This result was formulated in [BSW03] in a less general form, namely \( \Omega = \{0, 1\}^n \), \( C \) is the set of distributions with min-entropy at least \( k \), and \( D, D' \) are the classes of \([0, 1]\)-valued circuits of size \( s \) and \( O(s \cdot \frac{n}{\log n}) \) respectively. The inspection of the proof shows that: (a) the chosen space \( \Omega \) can be an arbitrary finite set, and the number \( n \) appearing in the assertion is equal to \( \log |\Omega| \), (b) the chosen set \( C \) can be replaced by an arbitrary convex set of distributions, (c) the complexity of the class \( D' \) is chosen only to ensure that \( D' \) contains all convex combinations of length \( O\left(\frac{\log |\Omega|}{\delta^2}\right) \) of elements of \( C \).

A.1 Separation of Metric and Smooth Entropy

Lemma 2 For any \( C \geq 1 \) and sufficiently large \( \ell \) there exists a boolean function \( f \) over \([0,1]^\ell\), such that \( \text{bias}(f) = 1 - 2^{-\sqrt{C}} \) and for all circuits \( D \) of size \( O\left(2^{\ell - \sqrt{C}}/\ell - \frac{C}{\epsilon} - 2\log(1/\delta)\right) \) we have

\[
\Pr_{x \leftarrow f^{-1}(\{1\})} (D(x) = f(x)) + \Pr_{x \leftarrow f^{-1}(\{0\})} (D(x) = f(x)) < 1 + \delta
\]

Proof. Chose a set \( A \) by sampling \( m = 2^\ell - C \) elements \( x \in \{0,1\}^\ell \) without replacement. The random variables \( D(x) \) for \( x \in A \) are not independent. However, the Hoeffding Inequality still holds for sampling without replacement and gives us

\[
\Pr_A \left( ED(U_A) - ED(U) > \frac{1}{2}\delta \right) \leq \exp \left( -\delta^2 2^{\ell - C} \right). \tag{36}
\]

Let \( B = A^c \). Since the set \( B \) can be viewed as chosen by sampling \( 2^\ell - 2^{\ell - C} \) elements from \( \{0,1\}^\ell \) without replacement, applying the Hoeffding Inequality again, we have

\[
\Pr_B \left( ED^c(U_B) - ED^c(U) > \frac{1}{2}\delta \right) \leq \exp \left( -\delta^2 2^{\ell} \left(1 - 2^{-C}\right)\right). \tag{37}
\]
Therefore, for every fixed circuit $D$ the inequality

$$\mathbf{E}D(U_A) + \mathbf{E}D^c(U_B) > 1 + \delta$$

holds with probability at most $2\exp\left(-2^C\delta^2\right)$ over choosing $A, B$. By a union bound over all $\exp(O(s \log s)) < \frac{1}{2} \exp\left(2^C\delta^2\right)$ circuits of size $s$, we obtain that there exists set $A$ and $B = A^c$ such that for every circuit $D$ of size $s$ we have

$$\mathbf{E}D(U_A) + \mathbf{E}D^c(U_B) \leq 1 + \delta$$

We define $f$ to be $1_A$ and the proof is finished.

**Remark 6.** If the assertion of the lemma is satisfied by a function $f$ then also by $1 - f$. Since $\text{bias}(f) = 1 - 2^{-C}$, replacing $f$ with $1 - f$ if necessary, we may assume that $\# \{x : f(x) = 1\} = (\frac{1}{2} - \frac{1}{2}\text{bias}(f))2^k = 2^C$. This in turn implies that for all circuits $D$

$$\Pr_x(f(x) = D(x)) =$$

$$= 2^{-C}\Pr_{x \leftarrow f^{-1}(1)}(D(x) = 1) + (1 - 2^{-C})\Pr_{x \leftarrow f^{-1}(0)}(D(x) = 0)$$

$$< 1 - 2^{-C} + \delta = \frac{1}{2} + \frac{1}{2}\text{bias}(f) + \delta.$$

Thus, we have retrieved the classical result on $\delta$-hard functions, as for every function $f$, the value of $f(x)$ can be guessed trivially (using a constant function) for at least $\frac{1}{2} + \frac{1}{2}\text{bias}(f)$ fraction of inputs $x$.

**Theorem 3** For sufficiently large $n$, and for any $C > 0$, $k < n - C$ and $\epsilon > 0$ there exists a pair of jointly distributed random variables $X \in \{0, 1\}^n$, $Z \in \{0, 1\}^m$ such that

(i) $\mathbf{H}^{1/2}_{\infty}(X|Z) \leq k + 1$

(ii) $\mathbf{H}^{\text{Metric}, \text{det}[0,1],s,\epsilon}(X|Z) \geq k + C$ for $s = \Omega(\frac{2^{k+m}/\epsilon^4}{(k+m) \log(2^{k+m}/\epsilon^4)})$

**Proof.** Fix a distribution $Z$ over $\{0, 1\}^m$. For every $z$, chose a $2^k$-element subsets $A(z)$ of $S = \{0, 1\}^{k+C}$. Let $B(z) = A(z)^c$. Let $X$ be a distribution (jointly distributed with $Z$) such that $X|Z = z$ is uniform over $A(z)$. We observe that $\mathbf{H}_{\infty}(X|Z) = k$ and $\mathbf{H}^{1/2}_{\infty}(X|Z) \leq k + 1$, since $\Delta(X|Z = z, Y_z) \leq 1/2$ for every distribution $Y_z$ over $\{0, 1\}^n$ such that $\mathbf{H}_{\infty}(Y_z) \geq k + 1$. Let $Y|Z = z$ be uniform over $S$. Assuming that $\mathbf{H}^{\text{Metric}, [0,1],s,\epsilon}_{\infty}(X|Z) < k + C$ with $\epsilon = (1 - 2^{-C})\delta$, from the definition of Metric Entropy (replacing $D$ with $D^c$ if necessary) we obtain for some $D$ of size $s$

$$\epsilon \leq \mathbf{E}D(X, Z) - \mathbf{E}D(U_S \times Z) =$$

$$= \mathbf{E}_{z \leftarrow Z}[\mathbf{E}D(X|Z = z, z) - \mathbf{E}D(U_S, z)]$$

$$= \mathbf{E}_{z \leftarrow Z}[\mathbf{E}D(U_{A(z)}, z) - \frac{|A(z)|}{|S|}\mathbf{E}D(U_{A(z)}) - \frac{|B(z)|}{|S|}\mathbf{E}D(U_{B(z)}, z, z)]$$

$$= \mathbf{E}_{z \leftarrow Z}[(1 - 2^{-C})\mathbf{E}D(U_{A(z)}, z) - (1 - 2^{-C})\mathbf{E}D(U_{B(z)}, z, z)]$$

$$= (1 - 2^{-C})\mathbf{E}_{z \leftarrow Z}[\mathbf{E}D(U_{A(z)}, z) + \mathbf{E}D^c(U_{A(z)}, z) - 1] \quad (38)$$

Therefore, for every distribution $Z$ there exists a circuit $D$ of size $s$ such that

$$\mathbf{E}_{z \leftarrow Z}\left[\mathbf{E}_{x \leftarrow A(z)}(D(x, z)) + \mathbf{E}_{x \leftarrow B(z)}(D^c(x, z))\right] \geq 1 + \delta \quad (39)$$

by a min-max theorem and obtain that there exists a circuit $D$ (not efficient itself but being a convex combination of circuits of size $s$) such that

$$\text{for all distributions } Z : \quad \mathbf{E}_{z \leftarrow Z}\left[\mathbf{E}_{x \leftarrow A(z)}(D(x, z)) + \mathbf{E}_{x \leftarrow B(z)}(D^c(x, z))\right] \geq 1 + \delta. \quad (40)$$
By a standard approximation via Chernoff Bounds, for some circuit of size \( s' = \mathcal{O}(\sqrt{k+m}s/\epsilon^2) \) we get

\[
\text{for all distributions } \mathbf{Z} : \quad \mathbf{E}_{z \sim \mathbf{Z}} \left[ \mathbf{E}_{x \sim A(z)} (D(x, z)) + \mathbf{E}_{x \sim B(z)} (D^c(x, z)) \right] \geq 1 + \delta/2. \tag{41}
\]

Especially, for every \( z \) we obtain

\[
\mathbf{E}_{x \sim A(z)} (D(x, z)) + \mathbf{E}_{x \sim B(z)} (D^c(x, z)) \geq 1 + \delta/2 \tag{42}
\]

Observe that this inequality is valid independently on the choice of \( A(z) \). We argue, that if \( A(z) \) are chosen at random, this inequality becomes a ‘hard task’ for small circuits. More precisely, we make use of the following lemma on hard functions

**Lemma 5.** Let \( \delta \in (0, 1), C > 0 \) and \( \mathcal{D} \) be a class of boolean randomized functions on \( \{0, 1\}^{\ell+m} \) of cardinality at most \( \exp \left( c \cdot 2^{\ell+m-C} \delta^2 \right) \) for universal constant \( c \). Then there exists a function \( f \) on \( \{0, 1\}^{\ell+m} \) such that bias\((f(\cdot, z)) = 1 - 2^{1-C} \) for every \( z \), with the following property: for every \( D \in \mathcal{D} \) there exists at least one \( z \) satisfying

\[
\Pr_{x : f(x, z) = 1} (D(x, z) = f(x, z)) + \Pr_{x : f(x, z) = 0} (D(x, z) = f(x, z)) < 1 + \delta. \tag{43}
\]

**Proof.** Fix a function \( D \in \mathcal{D} \). For every \( z \) chose a set \( A(z) \) by sampling \( m = 2^{\ell-C} \) elements \( x \in \{0, 1\}^\ell \) without replacement. The random variables \( D(x, z) \) for \( x \in A \) are not independent. However, the Hoeffding Inequality holds for sampling without replacement (see [Ser74] for instance) and gives us

\[
\Pr_{A(z)} (\mathbf{E}D(U_{A(z)}(z)) - \mathbf{E}D(U, z) \geq \delta/2) \leq \exp \left( -\Omega \left( \delta^2 2^{\ell-C} \right) \right) . \tag{44}
\]

Let \( B(z) = A(z)^c \). Since the set \( B(z) \) can be viewed as chosen by sampling \( 2^{\ell} - 2^{\ell-C} \) elements from \( \{0, 1\}^\ell \) without replacement, applying the Hoeffding Inequality again, we have

\[
\Pr_{B(z)} (\mathbf{E}D^c(U_{B(z)}(z)) - \mathbf{E}D^c(U, z) \geq \delta/2) \leq \exp \left( -\Omega \left( \delta^2 2^{\ell} (1 - 2^{-C}) \right) \right) . \tag{45}
\]

Since \( \mathbf{E}D(U) + \mathbf{E}D^c(U) = 1 \), inequalities (44) and (45) for every \( z \) yield

\[
\Pr_{A(z), B(z)} (\mathbf{E}D(U_{A(z)}(z)) + \mathbf{E}D^c(U_{B(z)}(z)) \leq 1 + \delta) \geq 1 - \exp \left( -\Omega \left( \delta^2 2^{\ell-C} \right) \right) . \tag{46}
\]

Thus, probability that all values \( z \) are ‘bad’ is equal to

\[
\Pr_{A, B} \left( \text{for every } z : \mathbf{E}D(U_{A(z)}(z)) + \mathbf{E}D^c(U_{B(z)}(z)) \leq 1 + \delta \right) \leq \exp \left( -\Omega \left( \delta^2 2^{\ell+m-C} \right) \right) , \tag{47}
\]

and by a union bound over all members of \( \mathcal{D} \) the result follows.

Note that condition bias\( f = 1 - 2^{1-C} \) in Lemma 5 means that either \( |f| = 2^C \) or \( |f^c| = 2^C \). Clearly, the lemma is valid also for \( f^c \). Thus, without losing generality, let \( |f| = 2^C \). Apply Lemma 5 to \( \ell = k + C \). Define the sets \( A(z) \) as \( A(z) = \{ x : f(x, z) = 1 \} \) and \( B(z) = \{ x : f^c(x, z) = 1 \} \). Since \( s' = \mathcal{O}(k+m)s/\epsilon^2 \), inequality (42) contradicts the lemma provided that

\[
\exp (s' \log s') \exp \left( -\Omega \left( 2^{k+m}\epsilon^2 \right) \right) < 1 \tag{48}
\]

or in other words if

\[
\frac{(k + m)s}{\epsilon^2} < c \cdot \frac{2^{k+m}\epsilon^2}{\log (2^{k+m}\epsilon^2)}, \tag{49}
\]

which is equivalent to

\[
s < c \cdot \frac{2^{k+m}\epsilon^4}{(k + m) \log (2^{k+m}\epsilon^2)} \tag{50}
\]
A.2 Characterizations of Rényi Metric Entropy

**Lemma 4** Let \( \alpha > 1 \) be fixed, let \( D : \{0, 1\}^n \to \{0, 1\} \) be a function and \( Y_k = \{ Y \in \{0, 1\}^n : H_\alpha(Y) \geq k \} \). Then

\[
\max_{Y \in Y_k} ED(Y) = \begin{cases} \frac{p_D}{|D|}, & \text{if } |D| < 2^k \\ 1, & \text{otherwise} \end{cases}
\]

where \( p_D, \) for \( |D| \leq 2^k, \) is the greatest number satisfying the following system

\[
\begin{cases} \frac{p_D}{|D|} + q_D_\alpha|D^c| = 2^{-(\alpha-1)k} \\ p_D|D| + q_D|D^c| = 1 \\ p_D, q_D \geq 0 \end{cases}
\]

Moreover, the solution \( p_D \) is unique provided that \( k < n - 1 \).

**Proof.** First we prove that \( \max_{Y \in Y_k} ED(Y) < 1 \) is equivalent to \( |D| < 2^k \). Suppose that \( \max_{Y \in Y_k} ED(Y) < 1 \). If \( |D^{-1}(1)| \geq 2^k \), then for \( Y \) being uniform over \( D^{-1}(1) \) we get a contradiction as \( ED(Y) = 1 \) and \( H_\alpha(Y) \geq H_\infty(Y) \geq k \). The other direction follows from the following Lemma, proved in the Appendix:

**Lemma 6.** Let \( X \in \{0, 1\}^n \) be a random variable satisfying \( H_\alpha(X) \geq k \). Then \( |\supp(P_X)| \geq 2^k \).

Assume that \( \max_{Y \in Y_k} ED(Y) < 1 \) and let \( Y' \) be a distribution maximizing \( ED(\cdot) \) over the set \( Y_k \). We will show, that \( Y \) may be assumed to be uniform if conditioned on the sets \( D^{-1}(0) \) and \( D^{-1}(1) \). The first part is clear because modifying the distribution \( Y \) outside the support of \( D \) we do not change the value \( ED(X) \). To prove the second one, define \( P_{Y'}(x) \) to be \( \sum_{x' \in D^{-1}(1)} P_Y(x') \) if \( x \in D^{-1}(1) \) and \( P_Y \) otherwise. By Jensen’s inequality we get

\[
\sum_{x \in D^{-1}(1)} P_{Y'}(x)^\alpha \leq \sum_{x \in D^{-1}(1)} P_Y(x)^\alpha \quad \text{and thus } H_\alpha(Y') \geq H_\alpha(Y).
\]

Since \( D \) is boolean, we also have \( ED(Y') = ED(Y) \). Therefore, for some \( p = p_D, q = q_D \) we have

\[
P_Y(x) = p1_{D^{-1}(1)}(x) + q1_{D^{-1}(0)}(x),
\]

where \( p, q \) should be chosen so that \( Y \) is a probability measure and satisfies the constraint \( H_\alpha(Y) \geq k \). These two conditions are exactly equations (2). Note that since the maximizer \( Y \in Y_k \) for \( D \) exists, this system certainly has a solution. To prove that this solution is unique, we observe that after substituting \( \gamma = p_D/|D| \) the first equation becomes \( f(\gamma) = 0 \) where \( f(\gamma) = \gamma^\alpha|D|^{1-\alpha} + (1 - \gamma)^\alpha|D^c|^{1-\alpha} - 2^{-k(\alpha-1)} \) and \( 0 \leq \gamma \leq 1/|D| \). Observe that the function \( f \) is strictly convex and, provided that \( k < n - 1 \), we have \( f(0) = (2^n - |D|)^{1-\alpha} - 2^{-k(\alpha-1)} < 0 \). Therefore, there can be at most one solution \( \gamma > 0 \).

A.3 Metric Min-Entropy Against Different Distinguishers

**Theorem 5** Let \( X \in \{0, 1\}^n \) and \( Z \in \{0, 1\}^n \) be random variables. Then we have \( H_\infty^{\text{det}, [0,1], s, \epsilon}(X|Z) = H_\infty^{\text{det}, [0,1], s', \epsilon}(X|Z) \) where \( s' \approx s \).

**Proof.** We need only to show that if \( H_\infty^{\text{det}, [0,1], s, \epsilon}(X|Z) \geq k \) then also \( H_\infty^{\text{det}, [0,1], s', \epsilon}(X|Z) \geq k \) for \( s' \approx s \). Let \( Y \) be the set of distributions of the random variables of the form \( (Y, Z) \) where \( Y \in \{0, 1\}^n \) and \( H_\infty(Y|Z) \geq k \). Suppose, that \( H_\infty^{\text{det}, [0,1], s', \epsilon}(X|Z) < k \). According to the definition, there exists a \( [0,1] \)-valued function \( D \) of complexity \( s' \) such that

\[
ED(X, Z) - \max_{Y \in Y} ED(Y, Z) \geq \epsilon
\]

We shall show that \( D \) can be replaced by a boolean distinguisher \( D' \) of (almost) the same complexity. Let \( P_{Y_0, Z} \) be a distribution that maximizes \( ED(\cdot) \) over \( Y \). It means that \( p_0 = P_{Y_0, Z} \) is a solution of the following constrained optimization problem in \( \mathbb{R}^{2^{n+m}} \):

maximize \[ \sum_{x,z} p(x, z) \cdot D(x, z) \]

s.t. \[ \sum_{x,z} p(x, z) = 1 \]

for every \( z \)

\[ -p(x, z) \leq 0, \quad \text{for all } x, z \]

\[ p(x, z) \leq 2^{-k}P_{Z}(z), \quad \text{for all } x, z \]
where the constraints in this optimization problem describe the set \( \mathcal{Y} \). We can assume that \( p_0 \) is chosen to be flat (for every \( x, z \) either \( p_0(x, z)/P_Z(z) = P_{Y \mid Z}(x) = 2^{-k} \) or \( p_0(x, z) = 0 \)) as otherwise we would have \( p_0 = tp_1 + (1-t)p_2 \) where \( p_1, p_2 \in \mathcal{Y} \) and then either \( p = p_1 \) or \( p = p_2 \) gives \( \langle D, p \rangle \geq \langle D, p_0 \rangle \). The proof will be complete, if we will find a function \( D' \) satisfying the following conditions:

(a) \( D' \) is boolean
(b) \( P_{Y_0, Z} \) is a maximizer for \( D' \) over \( \mathcal{Y} \), i.e. \( ED'(Y_0, Z) \geq ED'(Y, Z) \) for all \( P_{Y, Z} \in \mathcal{Y} \)
(c) \( ED'(X, Z) - ED'(Y_0, Z) \geq \epsilon \)
(d) \( D' \) has the complexity \( s \)

Consider now the condition in (b). It can be rewritten as \( \langle D', P_{Y_0, Z} - P_{Y, Z} \rangle \geq 0 \) for all \( P_{Y, Z} \in \mathcal{Y} \) (we indetify functions \( D', P_{X, Z}, P_{Y, Z} \) on \( \{0, 1\}^{2n+m} \) with vectors of \( \mathbb{R}^{2n+m} \)). The set of all such \( D' \in \mathbb{R}^{2n+m} \) is the normal cone of \( \mathcal{Y} \) at \( P_{Y_0, Z} \).

**Claim.** The normal cone of \( \mathcal{Y} \) at \( p_0 = P_{Y_0, Z} \), i.e. all real valued functions \( D' \) for (b), is described by the following condition: there exist the Lagrange Multipliers: \( \lambda^2(z), \lambda^3(x, z), \lambda^4(x, z) \geq 0 \) such that

\[
D'(x, z) = \lambda_2(z) - \lambda_3(x, z) + \lambda_4(x, z)
\]  

(56)

and satisfying the so called complementary slackness condition: \( \lambda_3(x, z), \lambda_4(x, z) \) can be nonzero only if the corresponding constraint is active, i.e. if \( p_0(x, z) = 0 \) or \( p_0(x, z) = 2^{-k}P_Z(z) \) respectively.

**Proof.** We can replace the first two (equality-type) constraints by the inequalities \( \sum_x p(x, z) \leq 1 \) and \( \sum_z p(x, z) \leq P_Z(z) \), as at the maximizer the equality will be achieved because of \( D'(x, z) \geq 0 \). Moreover, the first inequality can be dropped as it is implied by the second one. Now, the claim follows by standard facts from convex optimization: the normal cone of a set described by linear inequalities (a polyhedron) is a cone generated by the gradients of the ‘active’ constraints.

It is easy to see, that the above can be stated equivalently as follows:

**Claim.** The normal cone of \( \mathcal{Y} \) at \( P_{Y_0, Z} \), consists of all real valued functions \( D' \) satisfying

\[
D'(x_1, z) \geq D'(x_2, z) \quad \text{for every } z, x_1, x_2 \text{ such that } P_{Y_0, z = z}(x_1, z) = 2^{-k}, \ P_{Y_0, z = z}(x_2, z) = 0
\]  

(57)

The definition of \( P_{Y_0, Z} \) implies that \( D \) belongs to the normal cone of \( \mathcal{Y} \) at \( P_{Y_0, Z} \) and thus satisfies the assertion of Claim [A.3]. From Claim [A.3] it follows that also every threshold of \( D \); any function of the form \( D'(x, z) = \mathbf{1}_{\{D(x, z) > t\}} \) is also in the normal cone. Thus, every such \( D' \) satisfies [3, 5], and [4]. Finally, since have

\[
\epsilon \leq ED(X, Z) - ED(Y_0, Z) = \int_{t \in [0, 1]} (P_{X, Z}[D(X, Z) > t] - P_{X, Z}[D(X, Z) > t])
\]

(58)

for some \( t \) the corresponding function \( D' \) satisfies also [5]. This proves the first part of the theorem.

To prove the second one, suppose that there exists a \( [0, 1] \)-valued function \( D \) (possibly computationally inefficient) satisfying [4]. As in the proof of the first part, let \( P_{Y_0, Z} \) be a flat distribution maximizing \( ED(\cdot) \) over \( \mathcal{Y} \). Due to the first part of the theorem, we may assume that \( D \) is boolean. Suppose now, that \( D(x_0, z_0) = 0 \) for some \( (x_0, z_0) \notin \supp(X, Z) \). By [A.3], we obtain that \( D(x, z) = 0 \) for all \( (x, z) \in \supp(X, Z) \). But then we have \( ED(Y_0, Z) = 1 \) which contradicts to [5] as \( \epsilon > 0 \).

**Theorem 6** Let \( X \in \{0, 1\}^n \) and \( Z \in \{0, 1\}^m \) be random variables. Then we have \( H_{\text{Metric-rlx} \cdot \text{det}[0, 1], s, \epsilon} X \mid Z \) \( = \) \( H_{\text{Metric-rlx} \cdot \text{det}[0, 1], s', \epsilon} X \mid Z \) where \( s' \approx s \).

**Proof.** The proof follows easily by inspecting the previous proof for the case of the metric min-entropy. Namelly, for the relaxed definition we only need to remove the condition \( \sum_x p(x, z) = P_Z(z) \) from the description of the optimization problem given by equation [5].
A.4 Shannon Entropy against different distinguishers

Proof. Enumerate elements of \( \{0, 1\}^k \) by \( x_1, x_2, \ldots \) where \( x_d \) is the \( k \)-digit binary expansion of \( d - 1 \). We will construct the distribution \( X \) explicitly in the following way: for every \( d = 1, \ldots, 2^k \) let \( p = p(d) \) be a solution of (3) (we will prove later that this solution is unique). Define the sequence \( \gamma(d) = p(d) \cdot d \). Let \( X \) be a distribution on \( \{0, 1\}^k \) defined by \( P_X(x_1) = \gamma_1 \) and \( P_X(x_d) = \gamma(d) - \gamma(d-1) \). To prove that this construction works we need to show that \( X \) is a probability measure and satisfy claimed estimates on its entropy and pseudodistance. This task involves a lot of calculus to study the solutions of (3). The proof will be divided into subsequently claims.

Claim. Let \( n > 1 \) and \( k < n - 1 \). Then for every real number \( d \in [1, 2^k] \), the system (3) has a unique solution \( (p, q) = (p(d), q(d)) \). Moreover, for \( d < 2^k \), this solution is a smooth function of \( d \).

Proof. The proof will be split into three parts

The existence of a solution. First, we parametrize the solutions of the second equation of (3) by \( p(\gamma) = \frac{\gamma}{d} \) and \( q(\gamma) = \frac{1 - \gamma}{2^n - d} \) for \( \gamma \in [0, 1] \). Now, the left side of the first equation of (3), can be viewed as a function \( F \) of \( \gamma \). Namely, for fixed \( d \), define

\[
F(\gamma) = \gamma \log \gamma + (1 - \gamma) \log(1 - \gamma) - \gamma \log d - (1 - \gamma) \log(2^n - d)
\]

Then the system (3) is equivalent to the equation

\[
F(\gamma) = -k, \quad \gamma \in [0, 1]
\]

Observe that

\[
F(0) = -\log(2^n - d) < -k \leq -\log d = F(1)
\]

and therefore, by the Darboux Principle, we conclude that with some \( \gamma \in [0, 1] \) we have \( F(\gamma) = -k \). It follows that there exists numbers \( p, q \) being a solution of (3).

The uniqueness and smoothness. We calculate the derivative of \( F \) with respect to \( \gamma \):

\[
\frac{\partial F}{\partial \gamma} = \log \gamma - \log(1 - \gamma) + \log(2^n - d) - \log d
\]

\[
= \log \left( \frac{\gamma}{1 - \gamma} \right) - \log \left( \frac{d}{2^n - d} \right).
\]

Hence, the function \( F(\gamma) \) increases if \( \gamma > \frac{d}{2^n} \) and decreases for \( \gamma < \frac{d}{2^n} \). Since \( F(0) < -k \) there cannot be a solution \( \gamma \in (0, 1) \). Therefore, the solution \( p(\gamma), q(\gamma) \) exists only for some \( \gamma > \frac{d}{2^n} \) which satisfy \( F(\gamma) = -k \) and it is unique as the function \( F(\gamma) \) is then increasing. Finally, this number \( \gamma = \gamma(d) \) a \( C^\infty \) function of \( d \) by the Inverse Function Theorem, if only \( \gamma(d) < 1 \) or equivalently if \( d < 2^k \).

Claim. Let \( p(d), q(d) \) be the unique solution of the system (3). Define \( \gamma(d) = p(d) \cdot d \). Then

\[
\frac{\partial \gamma}{\partial d} = \frac{p - q}{\log p - \log q}
\]

and

\[
\frac{\partial^2 \gamma}{\partial d^2} = -\frac{d \left( \frac{\partial p}{\partial \gamma} \right)^2 + \frac{2^n - d}{q} \left( \frac{\partial q}{\partial \gamma} \right)^2}{\log p(d) - \log q(d)}
\]

Especially, \( \gamma(d) \) is a concave function.

Proof. For every \( d \) we have \( F(\gamma(d)) = -k \). Derivelating this equation with respect to \( d \), we obtain

\[
0 = \frac{\partial F(\gamma)}{\partial d} = \gamma' \log \gamma - \gamma' \log(1 - \gamma) + \gamma' \log(2^n - d) - \gamma' \log d - \gamma d + \frac{1 - \gamma}{2^n - d}
\]

\[
= \gamma' \left( \frac{\gamma}{d} \right) - \gamma' \log \left( \frac{1 - \gamma}{2^n - d} \right) - \left( \frac{\gamma}{d} - \frac{1 - \gamma}{2^n - d} \right)
\]

\[
= \gamma' \left( \frac{\gamma}{d} \right) - \gamma' \left( \frac{\gamma}{d} - \frac{1 - \gamma}{2^n - d} \right)
\]

\[
= \gamma' \left( \log p - \log q \right) - (p - q)
\]
From this we obtain the first identity. Taking the second derivative with respect to \(d\) we get

\[
0 = \gamma''(\log p - \log q) + \gamma'(p'/p - q'/q) - (p' - q')
\]

(69)

\[
= \gamma''(\log p - \log q) + d (p')^2 / p + (2^n - d) (q')^2 / q
\]

(70)

Clearly, \(\gamma'' < 0\).

Claim. For every \(d \in [1, 2^k]\) we have \(p > q\).

Proof. Suppose that \(p = q\) for some \(d\). Then \(p = q = 2^{-n}\) what contradicts to the first equation. Since \(p\) and \(q\) are continuous with respect to \(d\), we have either \(p > q\) or \(p < q\). The first holds for \(d = 2^k\).

Claim. Suppose that \(k < cn\) for some sufficiently small absolute constant \(c\). Then we have

\[
\gamma'(d) = \Theta \left( \frac{n - k}{d(n - \log d)} \right)
\]

(71)

Proof. Recall, that the number \(\gamma(d)\) is a solution of the equation \(F(\gamma(d)) = -k\) where \(F\) is a function defined by equation (59). This equation may be rewritten as

\[
\gamma = \frac{\log (2^n - d) - k + \mathbf{H}(\gamma)}{\log (2^n - d) - \log d}
\]

(72)

where \(\mathbf{H}(\gamma) = -\gamma \log \gamma - (1 - \gamma) \log (1 - \gamma)\) is the Shannon Entropy of a random variable taking two values with probabilities \(\gamma\) and \(1 - \gamma\) respectively. Since \(d \leq 2^k \leq 2^{n-2}\), we have the following estimates

\[
\log (2^n - d) - \log d \geq n - \log d - 1 \geq \frac{n - \log d}{2}
\]

(73)

\[
1 \geq 1 - 2^{-n} d \geq \frac{1}{2}
\]

(74)

Thus, by (73), (74) and the fact that \(\mathbf{H}(\gamma) \in [0, 1]\), we get

\[
\gamma(d) = \Theta \left( \frac{n - k}{n - \log d} \right)
\]

(75)

Differentiating with respect to \(d\) at a point \(d < 2^k\) we obtain

\[
\gamma'(d) = \log(e) \frac{(2^n - d) \log (2^n - d) + d \log(d) - k 2^n + 2^n \mathbf{H}(\gamma)}{d (2^n - d) (\log (2^n - d) - \log(d))^2} - \frac{\gamma'(d) \log \left( \frac{\gamma(d)}{1 - \gamma(d)} \right)}{\log (2^n - d) - \log(d)}
\]

(76)

From the inequalities (73) and (74) it follows that the first term in the expression above is equal to \(\Theta \left( \frac{n - k}{d(n - \log d)} \right)\). Now we will estimate the second term. Consider the case \(\gamma(d) < \frac{1}{2}\). Then

\[
\left| \frac{\log \left( \frac{\gamma(d)}{1 - \gamma(d)} \right)}{\log (2^n - d) - \log(d)} \right| \leq \left| \frac{\log \left( \frac{\gamma(1)}{1 - \gamma(1)} \right)}{\log (2^n - 2^k) - k} \right| = \log \left( \frac{n - k}{k} \right) + \Theta(1) = \Theta \left( \frac{\log n}{n - k} \right)
\]

(77)

where we have used the fact that \(\gamma(1) = 2^{-k} + \Theta \left( \frac{1}{k} \right)\) implied by (72), and the assumption \(k \leq n - 1\). If \(\gamma(d) > \frac{1}{2}\) then the second term is negative. Thus, provided that \(k < cn\) for sufficiently small constant \(c\), the result follows.

Claim. We have \(\mathbf{H}_{\text{indec.}}^{[0,1],0}(X) = k\).

Proof. Define for the convenience \(\gamma(0) = 0\). Observe, that the numbers \(\gamma(d)\) are increasing and since \(\gamma(2^k) = 1\) we have \(\sum_x \mathbf{P}_X(x) = 1\) (a telescopic sum). Therefore we have indeed defined a probability measure. Let \(D\) be any
boolean function on \{0, 1\}^n such that \(d = |D| < 2^k\). Since \(\gamma(d)\) is concave then \(P_X(x_0, x_1) = \gamma(d) - \gamma(d - 1)\) is decreasing with \(d\). Therefore

\[
\mathbf{E}D(X) = \sum_x P_X(x) \cdot D(x)
\]  
(78)

\[
\leq \max_{|I| = d} \sum_{i \in I} P_X(x_i)
\]  
(79)

\[
\leq \sum_{i=1}^d P_X(x_i)
\]  
(80)

\[= \gamma_d = p(d) \cdot |D|
\]  
(81)

and by the characterization in Lemma, the first part follows.

**Claim.** We have \(H_1(X) = \mathcal{O}(kn^{-1} + kn^{-1} \log n)\).

**Proof.** Now we estimate the entropy of \(X\). By definition

\[
H(X) = -\gamma(1) \log \gamma(1) - \sum_{d=2}^{2^k} (\gamma(d) - \gamma(d - 1)) \log (\gamma(d) - \gamma(d - 1))
\]  
(82)

Since \(\gamma\) is concave, we have \(\gamma(d) - \gamma(d - 1) \leq \gamma'(d - 1)\). The function \(t \rightarrow -t \log t\) is increasing for \(t \leq \frac{1}{2}\) and for \(d \geq 2\) and sufficiently large \(n\), by concavity again we have \(\gamma'(d - 1) \leq \gamma'(1) \leq \frac{1}{2}\). Hence,

\[
H(X) \leq -\gamma(1) \log \gamma(1) - \sum_{d=1}^{2^k-1} \gamma'(d) \log \gamma'(d)
\]  
(83)

The function \(d \rightarrow -\gamma'(d) \log \gamma'(d)\) is decreasing, as \(\gamma'(d)\) decreases and \(\gamma'(d) \leq \gamma'(1) \leq \frac{1}{2}\). Thus

\[
H(X) \leq \gamma(1) \log \gamma(1) - \gamma'(1) \log \gamma'(1) - \int_1^{2^k-1} \gamma'(d) \log \gamma'(d) \, dd
\]  
(84)

Using the estimate (71), for some constant \(C > 1\) we obtain

\[
- \int_1^{2^k} \gamma'(d) \log \gamma'(d) \, dd \leq - \int_1^{2^k} \frac{C(n-k)}{d(n-k)} \log \left(\frac{C(n-k)}{d(n-k)}\right) \, dd + \mathcal{O}(C)
\]  
(85)

\[
= - C \int_1^{2^k} \frac{n-k}{d(n-k)} \log \left(\frac{n-k}{d(n-k)}\right) \, dd + \mathcal{O}(C \log C)
\]  
(86)
Integrating and using the inequality $\ln(1 + x) \leq x$ for $x > -1$, we get

$$- \int_{1}^{2^k} \frac{n - k}{d(n - \log d)^2} \log \left( \frac{n - k}{d(n - \log d)^2} \right) \, dd = \ln 2 \cdot k + \ln^2 2 \cdot (n - k) \log \left( \frac{n - k}{n} \right) + \ln 2 \cdot \log(n - k)$$

$$+ \ln 2 \cdot \frac{n - k}{n} \log \frac{n - k}{n^2} + \frac{2k}{n}$$

$$= \ln 2 \left( k + (n - k) \ln \left( \frac{1 - k}{n} \right) \right) + \ln \left( 1 - \frac{k}{n} \right) + \frac{k}{n} \log n$$

$$+ \left( 1 - \frac{k}{n} \right) \log \left( 1 - \frac{k}{n} \right) + \frac{2k}{n}$$

$$\leq (1 + \ln 2) \cdot k^2 n^{-1} + kn^{-1} \log n \quad (87)$$

Finally, inequalities (84), (85) and (87) yield the result.

The proof follows by claims.

This result directly implies the following one:

**Corollary 10.** For some absolute constant $c > 0$ and every sufficiently large $n$ there exists a random variable such that $H^{\text{Metric}, \text{det}(0,1)}_1(X) = cn$ but $H_1(X) \leq cn/2$.

Now we give separation between randomized and deterministic distinguishers for the Shannon Entropy:

**Corollary 11.** For some absolute constant $c \in (0, 1)$, for every $n$ there exists a random variable $X \in \{0, 1\}^n$ such that:

(i) Metric Shannon Entropy of $X$ is $k \geq cn$, against all deterministic boolean functions and $\epsilon = 0$

(ii) Metric Shannon Entropy of $X$ is $k - \Omega(n)$, against all randomized circuits and $\epsilon = \Omega(1)$.

**Proof.** We will make use of the following result, which says that Shannon Entropy is continuous (almost Lipschitz) with respect to the statistical distance. The proof is technical and is given in the Appendix.

**Lemma 7.** Let $X, Y \in \{0, 1\}^n$ be random variables. Then $|H_1(X) - H_1(Y)| = O(n \Delta(X, Y)) - 2\Delta(X, Y) \log \Delta(X, Y)$.

**Corollary 12.** Let $X, Y \in \{0, 1\}^n$ be random variables such that $H_1(Y) - H_1(X) = d \geq 1$. Then

$$\Delta(X, Y) \geq \Omega(d/n)$$

**Proof.** Let $\epsilon = \Delta(X, Y)$. Lemma 7 gives us $|H_1(X) - H_1(Y)| < cne + 2\epsilon \log(1/\epsilon)$ for some absolute constant $c$. If $2 \log(1/\epsilon) > cn$ then $\epsilon \leq 2^{-cn/2}$ and for sufficiently large $n$ we get $1 \leq d \leq 4\epsilon \log(1/\epsilon) \leq 4 \cdot 2^{-cn/2} \cdot (cn/2) < 1$. Hence we must have $|H_1(X) - H_1(Y)| \leq 2cne$ for large $n$. For the remaining (finitely many) cases $n = 1, \ldots, N = N(c)$ for every $n$ we find a number $\gamma_n$ such that $|H_1(X) - H_1(Y)| \leq \gamma_n \Delta(X, Y)$, under the constraint $1 \leq |H_1(X) - H_1(Y)|$. By a compactness argument $\gamma_n$ are well defined and for the number $\gamma = \max(\gamma_1, \ldots, \gamma_N, 2c)$ we have $|H_1(X) - H_1(Y)| < \gamma \Delta(X, Y)$ for all $n$. Especially, $\Delta(X, Y) > \gamma^{-1} |H_1(X) - H_1(Y)|$, provided that $|H_1(X) - H_1(Y)| \geq 1$.

Let $X$ be distribution from Corollary 10. Consider the set $\mathcal{Y}$ of all distributions $Y \in \{0, 1\}^n$ with Shannon Entropy at least $3/4 cn$. By Corollary 12 we obtain that for every distribution $P_Y \in \mathcal{Y}$ there exists a $[0, 1]$-valued function $D$ such that $ED(X) - ED(Y) \geq \Omega(1)$. But it means that $H^{\text{HILL-det}[0,1], \Omega(1)}_1 \leq 3/4 cn$. Since there is no restriction on the complexity, the same holds for Metric entropy. Since for unbounded circuits, Metric Entropy against $[0, 1]$-valued and boolean randomized distinguishers is the same (up to a arbitrary small absolute error), the result follows.

**Lemma 7** Let $X \in \{0, 1\}^n$ be a random variable satisfying $H_\alpha(X) \geq k$. Then $|\text{supp}(P_X)| \geq 2^k$. 
Proof. Suppose that the distribution of $X$ is supported on some set $S$. The entropy constraint yields

$$
\sum_{x \in S} (P_X(x))^\alpha \leq 2^{-(\alpha-1)k}
$$

(88)
on the other side, the Jensen inequality gives us

$$
|S|^{-\alpha} = \left( \frac{1}{|S|} \sum_{x \in S} P_X(x) \right)^\alpha \leq \frac{1}{|S|} \sum_{x \in S} (P_X(x))^\alpha
$$

(89)
From these two inequalities it follows that $|S| \geq 2^k$.

Lemma 7.8 Let $X, Y \in \{0, 1\}^n$ be random variables. Then

$$
|H_1(X) - H_1(Y)| \leq O(n \Delta(X, Y) - 2 \Delta(X, Y) \log \Delta(X, Y)).
$$

Proof. Suppose that distributions of $X, Y$ are chosen to maximize $|H_1(X) - H_1(Y)|$ under the constraint with $\Delta(X, Y) = \epsilon$. Assume that $H_1(X) \leq k \leq H_1(Y)$. Consider the sets $S^- = \{x : P_Y(x) < P_X(x)\}$ and $S^+ = \{x : P_Y(x) > P_X(x)\}$. We can assume that they are nonempty as otherwise $P_X = P_Y$. The proof is divided into claims and starts with the following useful inequality:

Claim. Let $H(p, q) = -p \log p - q \log q$. Suppose that $0 \leq p \leq q$ and $p + q \leq 1$. Then

$$
H(p + \epsilon, q - \epsilon) < H(p, q), \quad -\epsilon < 0 \tag{90}
$$

$$
H(p + \epsilon, q - \epsilon) > H(p, q), \quad 0 < \epsilon < q - p \tag{91}
$$

$$
H(p + \epsilon, q - \epsilon) < H(p, q), \quad q - p < \epsilon \tag{92}
$$

Next we derive several properties of $P_X, P_Y$ over the sets $S^+, S^-$.

Claim. The set $S^-$ contains only one element $x = x'$.

Proof. Suppose that $S^-$ contains two points $x_1, x_2 \in S^-$, such that $P_X(x_1) < P_X(x_2)$. Consider a distribution $P_X'$ given by $P_{X'}(x_1) = P_X(x_1) - \delta$, $P_{X'}(x_2) = P_X(x_2) + \delta$ and $P_{X'}(x) = P_X(x)$ if $x \notin \{x_1, x_2\}$ where $\delta$ is sufficiently small positive number (from the definition of $S^-$ we have $P_X(x) \in (0, 1)$ for $x \in S^-$, provided that $S^-$ has at least two elements). Since

$$
-(a - \delta) \log(a - \delta) - (b + \delta) \log(b + \delta) < -a \log a - b \log b \quad \text{for} \quad 0 < a \leq b < 1, \quad \delta > 0 \tag{93}
$$

we have $H_1(X') < H_1(X)$. Since $\Delta(X, Y) = \Delta(X', Y)$ (for sufficiently small $\delta$) we get a contradiction with the choice of $X, Y$. Hence, we may assume that $|S^-| = 1$.

Claim. The distribution $P_Y$ is uniform over $S^+$.

Proof. We can assume $|S^+| > 1$. Suppose that that $P_Y(x_1) < P_Y(x_2)$ for $x_1, x_2 \in S^+$. Considering a distribution $P_Y'$ given by $P_{Y'}(x_1) = P_Y(x_1) + \delta$, $P_{Y'}(x_2) = P_Y(x_2) - \delta$ and $P_{Y'}(x) = P_Y(x)$ if $x \notin \{x_1, x_2\}$ (from the definition of $S^+$ we have $P_Y(x) \in (0, 1)$ for $x \in S^+$ provided that $|S^+| > 1$), by (93) we obtain $H_1(Y') > H_1(Y)$. Since $\Delta(X, Y') = \Delta(X, Y)$, we get a contradiction.

Claim. $P_X(x) > 0$ for at most one element $x = x'' \in S^+$.

Proof. Suppose that $0 < P_X(x_1) \leq P_X(x_2)$ for two different points $x_1, x_2 \in S^+$. Define $P_X'$ by $P_{X'}(x_1) = P_X(x_1) - \delta$, $P_{X'}(x_2) = P_X(x_2) + \delta$ and $P_{X'}(x) = P_X(x)$ if $x \notin \{x_1, x_2\}$ for sufficiently small $\delta > 0$. Then $\Delta(X', Y) = \Delta(X, Y)$ and by (93) we get $H_1(X') < H_1(X)$. Therefore, there is at most one point $x \in S^+$ such that $P_X(x) > 0$. Observe however, that this cannot hold: from the definition of $S^+$ we have $\epsilon = \sum_{x \in S^+} (P_Y(x) - P_X(x))$ and the fact that $Y$ is uniform over $S^+$ yields $\epsilon = |S^+| P_Y(x) - P_X(x)$. This implies $P_Y(x) = \frac{P_X(x) + \epsilon}{|S^+|}$. But now, the definition of $S^+$ yields the inequality $\epsilon > (|S^+| - 1) P_X(x)$ and then $P_Y(x) > 1$, a contradiction.
Claim. We have \( P_Y(x') = P_X(x') - \epsilon \) and \( P_Y(x'') = \frac{P_X(x'') + \epsilon}{|S^+|} \).

Proof. It is easy to see that \( \sum_{x \in S^-} (P_X(x) - P_Y(x)) = \sum_{x \in S^+} (P_Y(x) - P_X(x)) = \Delta(X, Y) \). This immediately implies the first equality. Second is obtained because of the previous two claims.

Claim. We have \( P_X(x') \geq P_X(x'') + \epsilon \).

Proof. Otherwise, we have \( P_X(x') < P_X(x'') + \epsilon \). Consider then a distribution \( P_X \) given by \( P_X' \): \( P_X'(x) = P_X(x') - \epsilon, P_X'(x') = P_X(x'') + \epsilon \) and \( P_X'(x) = P_X(x) \) if \( x \notin \{x', x''\} \). (This is a probability distribution because \( P_X'(x) = P_X(x') - \epsilon = P_Y(x') > 0 \). Then \( x = x'' \) is the only point such that \( P_X'(x) \geq P_Y(x) \). Thus \( \Delta(X', Y) = |P_X'(x') - P_Y(x'')| = |P_X(x'') - P_Y(x') + \epsilon| \). Observe now that the definition of \( S^+ \) implies \( P_X(x'') - P_Y(x') < 0 \) and \( \Delta(X', Y) \leq \epsilon \) implies \( -\epsilon < P_X(x'') - P_Y(x'') \). Therefore, \( \Delta(X', Y) \leq \epsilon \). Finally, note that \( P_X'(x') = P_X(x') - \epsilon < P_X(x'') \) and \( P_X'(x'') = P_X(x'') + \epsilon > P_X(x') \). If \( P_X(x'') \leq P_X(x') \), then by (90) we get \( H(P_X'(x'), P_X(x'')) < H(P_X(x''), P_X(x')) \). Otherwise, \( P_X(x') < P_X(x'') \) and since \( P_X'(x') < P_X(x') \) and \( P_X'(x'') \geq P_X(x'') \), (90) yields \( H(P_X'(x'), P_X'(x'')) < H(P_X(x'), P_X(x'')) \). Anyway, we obtain \( H_1(X') < H_1(X) \), a contradiction.

Now we are in position to give the final estimate. We consider two cases: \( \epsilon > 2^{-n} \) and \( \epsilon < 2^{-n} \).

Claim. Suppose that \( |S^+| \geq 2 \). Then \( P_X(x'') < \frac{\epsilon}{|S^+|} \).

Proof. The definition of \( S^+ \) implies that \( P_X(x'') < P_Y(x'') = \frac{P_X(x'') + \epsilon}{|S^+|} \).

Claim. For the case \( |S^+| > 1 \) we have \( H_1(Y) - H_1(X) \leq 6\epsilon + n\epsilon - \epsilon \log \epsilon \).

Proof. Consider the case \( |S^+| \geq 2 \). Define then \( P_X \) as \( P_X'(x'') = 0, P_X'(x') = P_X(x') + P_X(x''), P_X'(x) = P_X(x) \) if \( x \notin \{x', x''\} \). Note, that \( H_1(X') < H_1(X) \) by (90). Then we obtain
\[
H_1(Y) - H_1(X) \leq H_1(Y) - H_1(X')
\]
\[
= \sum_{x \in S^- \cup S^+} (P_X'(x) \log P_X'(x) - P_Y(x) \log P_Y(x)) = \sum_{x \in S^- \cup S^+} (P_X'(x) \log P_X'(x) - P_Y(x) \log P_Y(x) - \sum_{x \in S^+} P_Y(x) \log P_Y(x))
\]
\[
< P_X'(x') \log P_X'(x') - P_Y(x') \log P_Y(x') - \epsilon \log \epsilon + \epsilon \log |S^+|
\]
Since \( \Delta(X', Y) = \Delta(X, Y) + P_X(x'') \leq 2\epsilon \) and the function \( t \to t \log t \) is convex, it follows that
\[
P_X'(x') \log P_X'(x') - P_Y(x') \log P_Y(x') \leq |P_X'(x') - P_Y(x')| (t \to t \log t)\bigg|_{t=\log P_X'(x')}
\]
\[
\leq 2\epsilon \left( 1 / \ln 2 + \log P_X'(x') \right)
\]
\[
\leq 6\epsilon
\]
In turn, if $P_X(x'') < \epsilon$, then $P_Y(x'') = P_X(x'') + \epsilon < 2\epsilon$. Thus, provided that $\epsilon < 1/4$,

$$P_X(x'') \log(P_X(x'')) - P_Y(x'') \log P_Y(x'') < -P_Y(x'') \log P_Y(x'')$$

$$\leq -2\epsilon \log(2\epsilon)$$

(101) and if $\epsilon > 1/4$, then trivially $-P_Y(x'') \log P_Y(x'') \leq 1/2 < 2\epsilon$. Summing up, for the case $S^+$, we have proven that $H_1(Y) - H_1(X) \leq -2\epsilon \log \epsilon + 2\epsilon$.

We are left with the problem if estimating $H_1(Y) - H_1(X)$ for the extremely small values of $\epsilon$. 

