Pseudoentropy: Lower-bounds for Chain rules and Transformations

Krzysztof Pietrzak[⋆] and Maciej Skórski[⋆⋆]

IST Austria and University of Warsaw

Abstract. Computational notions of entropy have recently found many applications, including leakage-resilient cryptography, deterministic encryption or memory delegation. The two main types of results which make computational notions so useful are (1) Chain rules, which quantify by how much the computational entropy of a variable decreases if conditioned on some other variable (2) Transformations, which quantify to which extend one type of entropy implies another.

Such chain rules and transformations typically lose a significant amount in quality of the entropy, and are the reason why applying these results one gets rather weak quantitative security bounds. In this paper we for the first time prove lower bounds in this context, showing that existing results for transformations are, unfortunately, basically optimal for non-adaptive black-box reductions (and it's hard to imagine how non black-box reductions or adaptivity could be useful here.)

A variable X has k bits of HILL entropy of quality (ϵ, s) if there exists a variable Y with k bits min-entropy which cannot be distinguished from X with advantage ϵ by distinguishing circuits of size s. A weaker notion is Metric entropy, where we switch quantifiers, and only require that for every distinguisher of size s, such a Y exists.

We first describe our result concerning transformations. By definition, HILL implies Metric without any loss in quality. Metric entropy often comes up in applications, but must be transformed to HILL for meaningful security guarantees. The best known result states that if a variable X has k bits of Metric entropy of quality (ϵ, s), then it has k bits of HILL with quality $(2\epsilon, s\cdot\epsilon^2)$. We show that this loss of a factor $\Omega(\epsilon^{-2})$ in circuit size is necessary. In fact, we show the stronger result that this loss is already necessary when transforming so called deterministic real valued Metric entropy to randomised boolean Metric (both these variants of Metric entropy are implied by HILL without loss in quality).

The chain rule for HILL entropy states that if X has k bits of HILL entropy of quality (ϵ, s), then for any variable Z of length m, X conditioned on Z has $k-m$ bits of HILL entropy with quality $(\epsilon, s\cdot\epsilon^2/2^m)$. We show that a loss of $\Omega(2^m/\epsilon)$ in circuit size necessary here. Note that this still leaves a gap of ϵ between the known bound and our lower bound.

[⋆] Supported by the European Research Council consolidator grant (682815-TOCNeT).

[⋆⋆] Supported by the National Science Center, Poland (2015/17/N/ST6/03564)

1 Introduction

There exist various information theoretic notions of entropy that quantify the "uncertainty" of a random variable. A variable X has k bits of Shannon entropy if it cannot be compressed below k bits. In cryptography we mostly consider min-entropy, where we say that X has k bits of min-entropy, denoted $\mathbf{H}_\infty(X) = k$, if for any x, $\Pr[X = x] \leq 2^{-k}$.

In a cryptographic context, we often have to deal with variables that only appear to have high entropy to computationally bounded observers. The most important case is pseudorandomness, where we say that $X \in \{0,1\}^n$ is pseudorandom, if it cannot be distinguished from the uniform distribution over $\{0,1\}^n$.

More generally, we say that $X \in \{0,1\}^n$ has $k \leq n$ bits of HILL pseudoentropy [12], denoted $\mathbf{H}^{\mathsf{HILL}}_{\epsilon,s}(X) = k$ if it cannot be distinguished from some Y with $\mathbf{H}_\infty(Y) = k$ by any circuit of size s with advantage $> \epsilon$, note that we get pseudorandomness as a special case for $k = n$. We refer to k as the *quantity* and to (ϵ, s) as the *quality* of the entropy.

A weak notion of pseudoentropy called Metric pseudoentropy [3] often comes up in security proofs. This notion is defined like HILL, but with the quantifiers exchanged: We only require that for every distininguisher there exists a distribution $Y, \mathbf{H}_\infty(Y) = k$ that fools this particular distinguisher (not one such Y to fool them all).

HILL pseudoentropy is named after the authors of the [12] paper where it was introduced as a tool for constructing a pseudorandom generator from any one-way function. Their construction and analysis was subsequently improved in a series of works [11,13,28]. A lower bound on the number of calls to the underlying one-way function was given by [14].[1] More recently HILL pseudoentropy has been used in many other applications like leakage-resilient cryptography [6,17], deterministic encryption [7] and memory delegation [4].

The two most important types of tools we have to manipulate pseudoentropy are chain rules and transformations from one notion into another. Unfortunately, the known transformations and chain rules lose large factors in the quality of the entropy, which results in poor quantitative security bounds that can be achieved using these tools. In this paper we provide lower bounds, showing that unfortunately, the known results are tight (or almost tight for chain rules), at least when considering non-adaptive black-box reductions. Although black-box impossibility results have been overcome by non black-box constructions in the past [2], we find it hard to imagine how non black-box constructions or adaptivity could help in this setting. We believe that relative to the oracles we construct also adaptive reductions are impossible as adaptivity "obviously" is no of use, but proving this seems hard. Our results are summarized in Figures 1 and 2.

Complexity of the adversary. In order to prove a black-box separation, we will construct an oracle and prove the separation unconditionally relative to this

[1] Their $\Omega(n/log(n))$ lower bound matches existing constructions from *regular* one-way functions [10]. For general one-way functions this lower bound is still far of the best construction [28] making $\tilde{\Theta}(n^3)$ calls.

oracle, i.e., assuming all parties have access to it. This then shows that any construction/proof circumventing or separation in the plain model cannot be relativizing, which in particular rules out all black-box constructions [1,16].

In the discussion below we measure the complexity of adversaries only in terms of numbers of oracle queries. Of course, in the actual proof we also bound them in terms of circuit size. For our upper bounds the circuits will be of basically the same size as the number of oracle queries (so the number of oracle queries is a good indication of the actual size), whereas for the lower bounds, we can even consider circuits of exponential size, thus making the bounds stronger (basically, we just require that one cannot hard-code a large fraction of the function table of the oracle into the circuit).

Transformations. It is often easy to prove that a variable $X \in \{0,1\}^n$ has

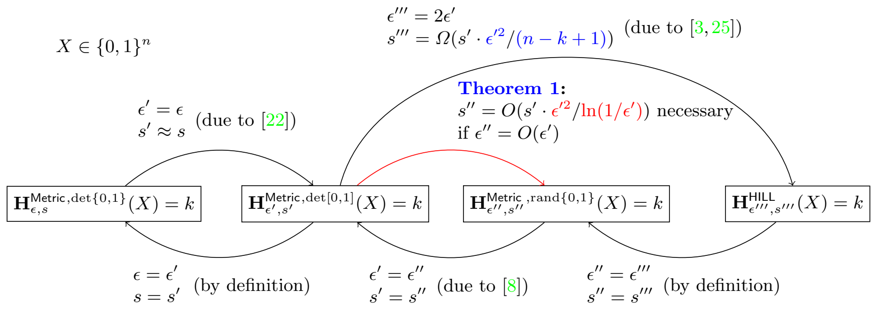
$X \in \{0,1\}^n$
$\epsilon''' = 2\epsilon'$
$s''' = \Omega(s' \cdot \epsilon'^2/(n-k+1))$ (due to [3,25])
Theorem 1:
$s'' = O(s' \cdot \epsilon'^2/\ln(1/\epsilon'))$ necessary
if $\epsilon'' = O(\epsilon')$
$\epsilon' = \epsilon$
$s' \approx s$ (due to [22])
$\mathbf{H}^{\mathsf{Metric},\det\{0,1\}}_{\epsilon,s}(X) = k$
$\mathbf{H}^{\mathsf{Metric},\det[0,1]}_{\epsilon',s'}(X) = k$
$\mathbf{H}^{\mathsf{Metric},\mathrm{rand}\{0,1\}}_{\epsilon'',s''}(X) = k$
$\mathbf{H}^{\mathsf{HILL}}_{\epsilon''',s'''}(X) = k$
$\epsilon = \epsilon'$
$s = s'$ (by definition)
$\epsilon' = \epsilon''$
$s' = s''$ (due to [8])
$\epsilon'' = \epsilon'''$
$s'' = s'''$ (by definition)

Fig. 1. Transformations: our bound comparing to the state of art. Our Thm. 1, stating that a loss of $\epsilon'^2/\ln(1/\epsilon')$ in circuit size is necessary for black-box reductions that show how deterministic implies randomized metric entropy (if the advantage ϵ' remains in the same order) requires $\epsilon' = 2^{-O(n-k+1)}$ and thus $\ln(1/\epsilon') \in O(n-k+1)$, so there's no contradiction between the transformations from [3,25] and our lower bound (i.e., the blue term is smaller than the red one).

so called Metric pseudoentropy against deterministic distinguishers, denoted $\mathbf{H}^{\mathsf{Metric},\det\{0,1\}}_{\epsilon,s}(X) = k$. Unfortunately, this notion is usually too weak to be useful, as it only states that for every (deterministic, boolean) distinguisher, there exists some Y with $\mathbf{H}_\infty(Y) = k$ that fools this particular distinguisher, but one usually needs a single Y that fools all (randomised) distinguishers, this is captured by HILL pseudoentropy.

Barak et al. [3] show that any variable $X \in \{0,1\}^n$ that has Metric entropy, also has the same amount of HILL entropy. Their proof uses the min-max theorem, and although it perseveres the amount k of entropy, the quality drops from (ϵ, s) to $(2\epsilon, \Omega(s \cdot \epsilon^2/n))$. A slightly better bound $\big(2\epsilon, \Omega(s \cdot \epsilon^2/(n+1-k))\big)$ (where again k is the amount of Metric entropy), was given recently in [25]. The argument uses the min-max theorem and some results on convex approximation in L_p spaces.

In Theorem 1 we show that this is optimal – up to a small factor $\Theta((n-k+1)/\ln(1/\epsilon))$ – as a loss of $\Omega(\ln(1/\epsilon)/\epsilon^2)$ in circuit size is necessary for any black-box reduction. Note that for sufficiently small $\epsilon \in 2^{-\Omega(n-k+1)}$ our bound even matches the positive result up to a small constant factor.

The high-level idea of our separation is as follows; We construct an oracle $\mathcal{O}$ and a variable $X \in \{0,1\}^n$, such that relative to this oracle X can be distinguished from any variable Y with high min-entropy when we can make one randomized query, but for any deterministic distinguisher A, we can find a Y with high min-entropy which A cannot distinguish from X.

To define $\mathcal{O}$, we first choose a uniformly random subset $S \in \{0,1\}^n$ of size $|S| = 2^m$. Moreover we chose a sufficiently large set of boolean functions $D_1(\cdot), \ldots, D_h(\cdot)$ as follows: for every $x \in S$ we set $D_i(x) = 1$ with probability $1/2$ and for every $x \notin S$, $D_i(x) = 1$ with probability $1/2 + \delta$.

Given any x, we can distinguish $x \in S$ from $x \notin S$ with advantage $\approx 2\delta$ by quering $D_i(x)$ *for a random* i. This shows that X cannot have much more than $\log(|S|) = m$ bits of HILL entropy (in fact, even probabilistic Metric entropy) as any variable Y with $\mathbf{H}_\infty(Y) \geqslant m+1$ has at least half of its support outside S, and thus can be distinguished with advantage $\approx 2\delta/2 = \delta$ with one query as just explained. Concretely (recall that in this informal discussion we measure size simply by the number of oracle queries)

$$\mathbf{H}^{\mathsf{Metric},\mathrm{rand}\{0,1\}}_{\delta,1}(X) \leqslant m+1$$

On the other hand, if the adversary is allowed q *deterministic* queries, then intuitively, the best he can do is to query $D_1(x), \ldots, D_q(x)$ and guess that $x \in S$ if less than a $1/2 + \delta/2$ fraction of the outputs is 1. But even if $q = 1/\delta^2$, this strategy will fail with constant probability. Thus, we can choose a Y with large support outside S (and thus also high min-entropy) which will fool this adversary. This shows that X does have large Metric entropy against deterministic distinguishers, even if we allow the adversaries to run in time $1/\delta^2$, concretely, we show that

$$\mathbf{H}^{\mathsf{Metric},\mathrm{det}\{0,1\}}_{\Theta(\delta),O(1/\delta^2)}(X) \geqslant n - O(\log(1/\delta))$$

The adversary. Let us stress that we show impossibility in the non-uniform setting, i.e., for any input length, the distinguisher circuit can depend arbitrarily on the oracle. Like in many non-uniform black-box separation results (including [19, 22, 24, 30, 31]), the type of adversaries for which we can rigorously prove the lower bound is not completely general, but the necessary restrictions seem "obviously" irrelevant. In particular, given some input x (where we must decide if $x \in S$), we only allow the adversary queries on input x. This doesn't seem like a real restriction as the distribution of $D_i(x')$ for any $x' \neq x$ is independent of x, and thus seems useless (but such queries can be used to make the success probability of the adversary on different inputs correlated, and this causes a problem in the proof). Moreover, we assume the adversary makes his queries non-adaptively, i.e., it choses the indices $i_1, \ldots, i_q$ before seeing the outputs of the queries $D_{i_1}(x), \ldots, D_{i_q}(x)$. As the distribution of all the D_i's is identical, this doesn't seem like a relevant restriction either.

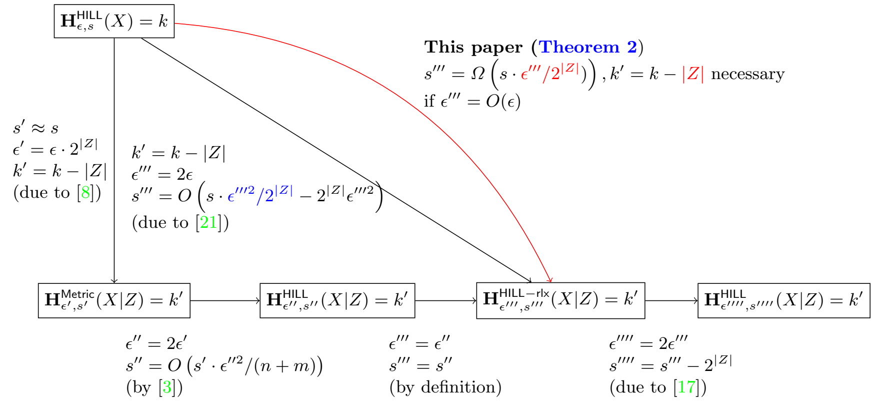
$\mathbf{H}^{\mathsf{HILL}}_{\epsilon,s}(X) = k$
This paper (Theorem 2)
$s''' = \Omega\left(s \cdot \epsilon'''/2^{|Z|}\right)\right)$, $k' = k - |Z|$ necessary
if $\epsilon''' = O(\epsilon)$
$s' \approx s$
$\epsilon' = \epsilon \cdot 2^{|Z|}$
$k' = k - |Z|$
(due to [8])
$k' = k - |Z|$
$\epsilon''' = 2\epsilon$
$s''' = O\left(s \cdot \epsilon'''^2/2^{|Z|} - 2^{|Z|}\epsilon'''^2\right)$
(due to [21])
$\mathbf{H}^{\mathsf{Metric}}_{\epsilon',s'}(X|Z) = k'$
$\mathbf{H}^{\mathsf{HILL}}_{\epsilon'',s''}(X|Z) = k'$
$\mathbf{H}^{\mathsf{HILL-rlx}}_{\epsilon''',s'''}(X|Z) = k'$
$\mathbf{H}^{\mathsf{HILL}}_{\epsilon'''',s''''}(X|Z) = k'$
$\epsilon'' = 2\epsilon'$
$s'' = O\left(s' \cdot \epsilon''^2/(n+m)\right)$
(by [3])
$\epsilon''' = \epsilon''$
$s''' = s''$
(by definition)
$\epsilon'''' = 2\epsilon'''$
$s'''' = s''' - 2^{|Z|}$
(due to [17])

Fig. 2. Chain Rules: our lower bounds comparing to the state of art. In the literature there are basically three approaches to prove a chain rule for HILL entropy. The first one reduces the problem to an efficient version of the dense model theorem [22], the second one uses the so called auxiliary input simulator [17], and the last one is by a convex optimization framework [21, 26]. The last approach yields a chain rule with a loss of $\approx 2^m/\epsilon^2$ in circuit size, where m is the length of leakage Z.

Chain Rules. Most (if not all) information theoretic entropy notions $H(.)$ satisfy some kind of chain rule, which states that the entropy of a variable X, when conditioned on another variable Z, can decrease by at most the bitlength $|Z|$ of Z, i.e., $H(X|Z) \geqslant H(X) - |Z|$.

Such a chain rule also holds for some computational notions of entropy. For HILL entropy a chain rule was first proven in [6, 22] by a variant of the *dense model theorem*, and was improved by Fuller and Reyzin [8]. A different approach using a *simulator* was proposed in [17] and later improved by Vadhan and Zheng [29]. A unified approach, based on convex optimization techniques was proposed recently in [21, 26] achieving best bounds so far.

The "dense model theorem approach" [8] proceeds as follows: one shows that if X has k bits of HILL entropy, then $X|Z$ has $k-m$ (where $Z \in \{0,1\}^m$) bits of Metric entropy. In a second step one applies a Metric to HILL transformation, first proven by Barak et al. [3], to argue that $X|Z$ has also large HILL. The first step loses a factor 2^m in advantage, the second another $2^{2m}\epsilon^2$ in circuit size. Eventually, the loss in circuit size is $2^{2m}/\epsilon^2$ and the loss in advantage is 2^m which measured in terms of the security ratio size/advantage gives a total loss of $2^m/\epsilon^2$.

A more direct "simulator" approach [29] loses only a multiplicative factor $2^m/\epsilon^2$ in circuit size (there's also an additive $1/\epsilon^2$ term) but there is no loss in advantage. The additive term can be improved to only $2^m\epsilon^2$ as shown in [21,26].

In this paper we show that a loss of $2^m/\epsilon$ is necessary. Note that this still is a factor $1/\epsilon$ away from the positive result. Our result as stated in Theorem 2 is a bit stronger as just outlined, as we show that the loss is necessary even if we only want a bound on the "relaxed" HILL entropy of $X|Z$ (a notion weaker than standard HILL).

To prove our lower bound, we construct an oracle $\mathcal{O}(.)$, together with a joint distribution $(X, Z) \in \{0,1\}^n \times \{0,1\}^m$. We want X to have high HILL entropy relative to $\mathcal{O}(.)$, but when conditioning on Z it should decrease as much as possible (in quantity and quality).

We first consider the case $m = 1$, i.e., the conditional part Z is just one bit. For $n \gg \ell \gg m = 1$ the oracle $\mathcal{O}(.)$ and the distribution (X, Z) is defined as follows. We sample (once and for all) two (disjoint) random subset $\mathcal{X}_0, \mathcal{X}_1 \subseteq \{0,1\}^n$ of size $|\mathcal{X}_0| = |\mathcal{X}_1| = 2^{\ell-1}$, let $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$. The oracle $\mathcal{O}(.)$ on input x is defined as follows (below B_p denotes the Bernoulli distribution with parameter p, i.e., $\Pr[b = 1 \ : \ b \leftarrow B_p] = p$).

- If $x \in \mathcal{X}_0$ output a sample of $B_{1/2+\delta}$.
- If $x \in \mathcal{X}_1$ output a sample of $B_{1/2-\delta}$.
- Otherwise, if $x \notin \mathcal{X}$, output a sample of $B_{1/2}$.

Note that our oracle $\mathcal{O}(.)$ is probabilistic, but it can be "derandomized" as we'll explain at the beginning of Section 4. The joint distribution (X, Z) is sampled by first sampling a random bit $Z \leftarrow \{0,1\}$ and then $X \leftarrow \mathcal{X}_Z$.

Given a tuple (V, Z), we can distinguish the case $V = X$ from the case where $V = Y$ for any Y with large support outside of $\mathcal{X}$ (X has min-entropy ℓ, so let's say we take a variable Y with $\mathbf{H}_\infty(Y|Z) \geqslant \ell + 1$ which will have at least half of its support outside $\mathcal{X}$) with advantage $\Theta(\delta)$ by quering $\alpha \leftarrow \mathcal{O}(V, Z)$, and outputting $\beta = \alpha \oplus Z$.

- If $(V, Z) = (X, Z)$ then $\Pr[\beta = 1] = 1/2 + \delta$. To see this, consider the case $Z = 0$, then $\Pr[\beta = 1] = \Pr[\alpha = 1] = \Pr[\mathcal{O}(X) = 1] = 1/2 + \delta$.
- If $(V, Z) = (Y, Z)$ then $\Pr[\beta = 1] = \Pr[Y \notin \mathcal{X}](1/2) + \Pr[Y \in \mathcal{X}](1/2 + \delta) \leq 1/2 + \delta/2$.

Therefore $X|Z$ doesn't have $\ell + 1$ bits of HILL entropy

$$\mathbf{H}^{\mathsf{HILL}}_{\delta/2,1}(X|Z) < \ell + 1$$

On the other hand, we claim that X (without Z but access to $\mathcal{O}(.)$) cannot be distinguished from the uniform distribution over $\{0,1\}^n$ with advantage $\Theta(\delta)$ unless we allow the distinguisher $\Omega(1/\delta)$ oracle queries (the hidden constant in $\Theta(\delta)$ can be made arbitrary large by stetting the hidden constant in $\Omega(1/\delta)$ small enough), i.e.,

$$\mathbf{H}^{\mathsf{HILL}}_{\Theta(\delta),\Omega(1/\delta)}(X) = n \tag{1}$$

To see why (1) holds, we first note that given some V, a single oracle query is useless to tell whether $V = X$ or $V = U_n$: although in the case where $V = X \in \mathcal{X}_Z$ the output $\mathcal{O}(X)$ will have bias δ, one can't decide in which direction the bias goes as Z is (unconditionally) pseudorandom. If we're allowed in the order $1/\delta^2$ queries, we can distinguish X from U_n with constant advantage, as with $1/\delta^2$ samples one can distinguish the distribution $B_{1/2+\delta}$ (or $B_{1/2-\delta}$) from $B_{1/2}$ with constant advantage. If we just want $\Theta(\delta)$ advantage, $\Omega(1/\delta)$ samples are necessary, which proves (1). While it is easy to prove that for the coin with bias δ one needs $O\left(1/\delta^2\right)$ trials to achieve 99% of certainty, finding the number of trials for some confidence level in $o(1)$ as in our case, is more challenging. We solve this problem by a tricky application of *Renyi divergences*[2] The statement of our "coin problem" with precise bounds is given in Lemma 3.

So far, we have only sketched the case $m = 1$. For $m > 1$, we define a random function $\pi : \{0,1\}^n \to \{0,1\}^{m-1}$. The oracle now takes an extra $m-1$ bit string j, and for $x \in \mathcal{X}$, the output of $\mathcal{O}(x, j)$ only has bias δ if $\pi(x) = j$ (and outputs a uniform bit everywhere else). We define the joint distribution (X, Z) by sampling $X \leftarrow \mathcal{X}$, define Z' s.t. $X \in \mathcal{X}_{Z'}$, and set $Z = \pi(X) \| Z'$. Now, given Z, we can make one query $\alpha \leftarrow \mathcal{O}(V, Z[1 \ldots m-1])$ and output $\beta = \alpha \oplus Z[m]$, where, as before, getting advantage δ in distinguishing X from any Y with min-entropy $\geq \ell + 1$.

On the other hand, given some V (but no Z) it is now even harder to tell if $V = X$ or $V = Y$. Not only don't we know in which direction the bias goes as before in the case $m = 1$ (this information is encoded in the last bit $Z[m]$ of Z), but we also don't know on which index $\pi(V)$ (in the case $V = X$) we have to query the oracle to observe any bias at all. As there are 2^{m-1} possible choices for $\pi(V)$, this intuitively means we need 2^{m-1} times as many samples as before to observe any bias, which generalises (1) to

$$\mathbf{H}^{\mathsf{HILL}}_{\Theta(\delta),\Omega(2^{m-1}/\delta)}(X) = n$$

1.1 Some implications of our lower bounds

Leakage Resilient Cryptography. The chain rule for HILL entropy is a main technical tool used in several security proofs like the construction of leakage-resilient schemes [6,20]. Here, the quantitative bound provided by the chain rule directly translates into the amount of leakage these constructions can tolerate. Our Theorem 2 implies a lower bound on the necessary security degradation for this proof technique. This degradation is, unfortunately, rather severe: even if we just leak $m = 1$ bit, we will lose a factor $2^m/\epsilon$, which for a typical security parameter $\epsilon = 2^{-80}$ means a security degradation of "80 bits".

[2] Lower bounds [30,31] also require nontrivial binomial estimates. They were obtained, however by direct and involved calculations.

Let us also mention that Theorem 2 answers a question raised by Fuller and Reyzin [8], showing that for any chain rule the *simultaneous loss* in quality and quantity is necessary,[3]

Faking Auxiliary Inputs. [17, 27, 29] consider the question how efficiently one can "fake" auxiliary inputs. Concretely, given any joint distribution (X, Z) with $Z \in \{0,1\}^m$, construct an *efficient* simulator h s.t. $(X, h(X))$ is (ϵ, s)-indistinguishable from (X, Z). For example [29] gives a simulator h of complexity $O\left(2^m\epsilon^2 \cdot s\right)$ (plus additive terms independent of s). This result has found many applications in leakage-resilient crypto, complexity theory and zero-knowledge theory. The best known lower bound (assuming exponentially hard OWFs) is $\Omega\left(\max(2^m, 1/\epsilon)\right)$. Since the chain rule for relaxed HILL entropy follows by a simulator argument [17] with the same complexity loss, our Theorem 2 yields a better lower bound $\Omega\left(2^m/\epsilon\right)$ on the complexity of simulating auxiliary inputs.

Dense Model Theorem. The computational dense model theorem [22] says, roughly speaking, that dense subsets of pseudorandom distributions are computationally indistinguishable from true dense distributions. It has found applications including differential privacy, memory delegation, graph decompositions and additive combinatorics. It is well known that the worst-case chain rule for HILL-entropy is equivalent to the dense model theorem, as one can think of dense distributions as uniform distributions X given short leakage Z. For settings with constant density, which correspond to $|Z| = O\left(1\right)$, HILL and relaxed HILL entropy are equivalent [17]; moreover, the complexity loss in the chain rule is then equal to the cost of transforming Metric Entropy into HILL Entropy. Now our Theorem 1 implies a necessary loss in circuit size $\Omega\left(1/\epsilon^2\right)$ if one wants ϵ-indistinguishability. This way we reprove the tight lower bound due to Zhang [31] for constant densities.

2 Basic Definitions

Let X_1 and X_2 be two distributions over the same finite set. The *statistical distance* of X_1 and X_2 equals $\mathrm{SD}\left(X_1; X_2\right) = \frac{1}{2}\sum_x \left|\Pr[X_1 = x] - \Pr[X_2 = x]\right|$.

Definition 1 (Min-Entropy). *A random variable X has* min-entropy k*, denoted by* $\mathbf{H}_\infty\left(X\right) = k$*, if* $\max_x \Pr[X = x] \leqslant 2^{-k}$.

Definition 2 (Average conditional min-Entropy [5]). *For a pair (X, Z) of random variables, the* average min-entropy *of X conditioned on Z is*

$$\widetilde{\mathbf{H}}_\infty(X|Z) = -\log \mathop{\mathbb{E}}_{z\leftarrow Z}[\max_x \Pr[X = x|Z = z]] = -\log \mathop{\mathbb{E}}_{z\leftarrow Z}[2^{-\mathbf{H}_\infty(X|Z=z)}]$$

[3] Their question was about chain rules bounding the worst-case entropy, that is bounding $\mathbf{H}^{\mathsf{HILL}}(X|Z = z)$ for every z. Our result, stated simply for average entropy $\mathbf{H}^{\mathsf{HILL}}(X|Z)$, is much more general and applies to qualitatively better chain rules obtained by simulator arguments.

Distinguishers. We consider several classes of distinguishers. With $\mathcal{D}_s^{\mathsf{rand},\{0,1\}}$ we denote the class of randomized circuits of size at most s with boolean output (this is the standard non-uniform class of distinguishers considered in cryptographic definitions). The class $\mathcal{D}_s^{\mathsf{rand},[0,1]}$ is defined analogously, but with real valued output in $[0,1]$. $\mathcal{D}_s^{\mathsf{det},\{0,1\}}, \mathcal{D}_s^{\mathsf{det},[0,1]}$ are defined as the corresponding classes for *deterministic* circuits. With $\Delta^D(X;Y) = |\,\mathbb{E}_X[D(X)] - \mathbb{E}_Y[D(Y)]$ we denote D's advantage in distinguishing X and Y.

Definition 3 (HILL pseudoentropy [12, 15]). *A variable X has* HILL entropy *at least k if*

$$\mathbf{H}^{\mathsf{HILL}}_{\epsilon,s}(X) \geq k \iff \exists Y\ ,\ \mathbf{H}_\infty(Y) = k\ \forall D \in \mathcal{D}_s^{\mathsf{rand},\{0,1\}}\ :\ \Delta^D(X;Y) \leq \epsilon$$

For a joint distribution (X,Z), we say that X has k bits conditonal Hill entropy *(conditionned on Z) if*

$$\begin{aligned}&\mathbf{H}^{\mathsf{HILL}}_{\epsilon,s}(X|Z) \geq k\\ \iff &\exists (Y,Z), \widetilde{\mathbf{H}}_\infty(Y|Z) = k\ \forall D \in \mathcal{D}_s^{\mathsf{rand},\{0,1\}} : \Delta^D((X,Z);(Y,Z)) \leq \epsilon\end{aligned}$$

Definition 4 (Metric pseudoentropy [3]). *A variable X has* Metric entropy *at least k if*

$$\mathbf{H}^{\mathsf{Metric}}_{\epsilon,s}(X) \geq k \iff \forall D \in \mathcal{D}_s^{\mathsf{rand},\{0,1\}} \exists Y_D\ ,\ \mathbf{H}_\infty(Y_D) = k\ :\ \Delta^D(X;Y_D) \leq \epsilon$$

Metric star entropy *is defined analogousely but using deterministic real valued distinguishers*

$$\mathbf{H}^{\mathsf{Metric*}}_{\epsilon,s}(X) \geq k \iff \forall D \in \mathcal{D}_s^{\mathsf{det},[0,1]} \exists Y_D\ ,\ \mathbf{H}_\infty(Y_D) = k\ :\ \Delta^D(X;Y_D) \leq \epsilon$$

Relaxed versions of HILL and Metric entropy. A weaker notion of conditional HILL entropy allows the conditional part to be replaced by some computationally indistinguishable variable

Definition 5 (Relaxed HILL pseudoentropy [9, 23]). *For a joint distribution (X,Z) we say that X has* relaxed HILL entropy *k conditioned on Z if*

$$\begin{aligned}&\mathbf{H}^{\mathsf{HILL-rlx}}_{\epsilon,s}(X|Z) \geq k\\ \iff &\exists (Y,Z'), \widetilde{\mathbf{H}}_\infty(Y|Z') = k, \forall D \in \mathcal{D}_s^{\mathsf{rand},\{0,1\}}\ ,\ :\ \Delta^D((X,Z);(Y,Z')) \leq \epsilon\end{aligned}$$

The above notion of *relaxed* HILL satisfies a chain rule whereas the chain rule for the standard definition of conditional HILL entropy is known to be false [18]. One can analogously define relaxed variants of metric entropy, we won't give these as they will not be required in this paper.

Pseudoentropy against different distinguisher classes. For randomized distinguishers, it's irrelevant if the output is boolean or real values, as we can replace any $D \in \mathcal{D}_s^{\mathsf{rand},[0,1]}$ with a $D' \in \mathcal{D}^{\mathsf{rand},\{0,1\}}$ s.t. $\mathbb{E}[D'(X)] = \mathbb{E}[D(X)]$ by setting (for any x) $\Pr[D'(x) = 1] = \mathbb{E}[D(x)]$. For HILL entropy (as well as for its relaxed version), it also doesn't matter if we consider randomized or deterministic distinguishers in Definition 3, as we always can "fix" the randomness to an optimal value. This is no longer true for metric entropy,[4] and thus the distinction between metric and metric star entropy is crucial.

3 A Lower Bound on Metric-to-HILL Transformations

Theorem 1. *For every n, k, m and ϵ such that $n \geqslant k + \log(1/\epsilon) + 4$, $\frac{1}{8} > \epsilon$ and $n - 1 \geq m > 6\log(1/\epsilon)$ there exist an oracle $\mathcal{O}$ and a distribution X over $\{0,1\}^n$ such that*

$$\mathbf{H}^{\mathsf{Metric},\det\{0,1\}}_{\epsilon,T}(X) \geqslant k \tag{2}$$

here the complexity T denotes any circuit of size $2^{O(m)}$ that makes at most $\frac{\ln(2/\epsilon)}{216\epsilon^2}$ non-adaptive queries and, simultaneously,

$$\mathbf{H}^{\mathsf{Metric},\mathrm{rand}\{0,1\}}_{2\epsilon,T'}(X) \leqslant m + 1 \tag{3}$$

where the distinguishers size T' is only $O(n)$ and the query complexity is 1.

Let S be a random subset of $\{0,1\}^n$ of size 2^m, where $m \leqslant n - 1$, and let $D_1, \ldots, D_h$ be boolean functions drawn independently from the following distribution D: $D(x) = 1$ on S with probability p if $x \in S$ and $D(x) = 1$ with probability q if $x \in S^c$, where $p > q$ and $p + q = 1$. Denote $X = U_S$. We will argue that the metric entropy against a probabilistic adversary who is allowed one query is roughly m with advantage $\Omega(p - q)$. But the metric entropy against non-adaptive deterministic adversary who can make t queries of the form $D_i(x)$ is much bigger, even if $t = O\left((p - q)^{-2}\right)$. Let us sketch an informal argument before we give the actual proof. We need to prove two facts:

(i) There is a probabilistic adversary A^* such that with high probability over $X, D_1, \ldots, D_h$ we have $\Delta^{\mathsf{A}^*}(X, Y) = \Omega(p - q)$ for all Y with $\mathbf{H}_\infty(Y) \geqslant m + 1$.

(ii) For every deterministic adversary A making at most $t = O\left((p - q)^{-2}\right)$ non-adaptive queries, with high probability over $X, D_1, \ldots, D_h$ we have $\Delta^{\mathsf{A}}(X; Y) = 0$ for some Y with $\mathbf{H}_\infty(Y) = n - \Theta(1)$.

To prove (i) we observe that the probabilistic adversary can distinugish between S and S^c by comparing the bias of ones. We simply let A^* forward its input to D_i for a randomly chosen i, i.e.,

$$\mathsf{A}^*(x) = D_i(x), \quad i \leftarrow [1, \ldots, h]$$

With extremely high probability we have $\Pr[\mathsf{A}^*(x) = 1] \in [p - \delta, p + \delta]$ if $x \in S$ and $\Pr[\mathsf{A}^*(x) = 1] \in [q - \delta, q + \delta]$ if $x \notin S$ for some $\delta \ll p - q$ (by a Chernoff

[4] It might be hard to find a high min-entropy distribution Y that fools a randomized distinguisher D, but this task can become easy once D's randomness is fixed.

bound, δ drops exponentially fast in h, so we just have to set h large enough). We have then $\Pr[\mathsf{A}^*(X) = 1] \geqslant p + \delta$ and $\Pr[\mathsf{A}^*(Y) = 1] \leqslant 1/2 \cdot (p + q + 2\delta)$ for every Y of min-entropy at least $m + 1$ (since then $\Pr[Y \in S] \leqslant 1/2$). This yields $\Delta^{\mathsf{A}^*}(X; Y) = (p - q)/2$. In order to prove (ii) one might intuitively argue that the best a t-query deterministic adversary can do to contradict to (ii), is to guess whether some value x has bias p or $q = 1 - p$, by taking the majority of t samples

$$\mathsf{A}(x) = \mathrm{Maj}(D_1(x), \ldots, D_t(x))$$

But even if $t = \Theta(1/(p-q)^2)$, majority will fail to predict the bias with constant probability. This means there exists a variable Y with min-entropy $n - \Theta(1)$ such that $\Pr[\mathsf{A}(Y) = 1] = \Pr[\mathsf{A}(X) = 1]$. The full proof gives quantitative forms of (i) and (ii), showing essentially that "majority is best" and appears in Appendix A.

4 Lower Bounds on Chain Rules

For any $n \gg \ell \gg m$, we construct a distribution $(X, Z) \in \{0,1\}^n \times \{0,1\}^m$ and an oracle $\mathcal{O}(.)$ such that relative to this oracle, X has very large HILL entropy but the HILL entropy of $X|Z$ is much lower in quantity and quality: for arbitrary $n \gg \ell \gg m$ (where $|Z| = m$, $X \in \{0,1\}^n$), the quantity drops from n to $\ell - m + 2$ (it particular, by much more than $|Z| = m$), even if we allow for a $2^m/\epsilon$ drop in quality.

Theorem 2 (A lower bound on the chain rule for $\mathbf{H}^{\mathsf{HILL-rlx}}$). *There exists a joint distribution (X, Z) over $\{0,1\}^n \times \{0,1\}^m$, and an oracle $\mathcal{O}$ such that, relative to $\mathcal{O}$, for any (ℓ, δ) such that $\frac{n}{2} - \frac{\log(1/\delta)}{2} > m$ and $\ell > m + 6\log(1/\delta)$, we have*

$$\mathbf{H}^{\mathsf{HILL}}_{\delta/2,T}(X) = n \tag{4}$$

where[5] *$T > c \cdot 2^m/\delta$ with some absolute constant c but*

$$\mathbf{H}^{\mathsf{HILL-rlx}}_{\delta/2,T'}(X|Z) < \ell + 1 \tag{5}$$

where T' captures a circuit of size only $O(n)$ making only 1 oracle query.

Remark 1 (On the technical restrictions). Note that the assumptions on ℓ and δ are automatically satisfied in most interesting settings, as typically we assume $m \ll n$ and $\log(1/\delta) \ll n$.

Remark 2 (A strict separation). The theorem also holds if we insist on a larger distinguishing advantage after leakage. Concretely, allowing for more than just one oracle query, the $\delta/2$ advantage in (5) can be amplified to $C\delta$ for any constant C assuming δ is small enough to start with (see Remark 4 in the proof).

[5] The class of adversaries here consists of all circuits with the total number of gates, including oracle gates, at most T. Theorem 2 is also true when the circuit size s is much bigger than the total number of oracle gates T (under some assumption on s, ℓ, ϵ). For simplicity, we do not state this version.

The full proof appears in Appendix B. The heart of the argument is a lower bound on the query complexity for the corresponding "coin problem": we need to distinguish between T random bits, and the distribution where we sample equally likely T independent bits B_p or T independent bits B_q where $p = \frac{1}{2} + \delta$ and $q = 1 - p$. (see Appendix C for more details). The rest of the proof is based on a standard concentration argument, using extensively Chernoff Bounds.

5 Open Problems

As shown in Figure 2, there remains a gap between the best proofs for the chain-rule, which lose a factor $\epsilon^2/2^{|Z|}$ in circuit size, and the required loss of $\epsilon/2^{|Z|}$ we prove in this paper. Closing this bound by either improving the proof for the chain-rule or give an improved lower bound remains an intriguing open problem.

Our lower bounds are only proven for adversaries that make their queries non-adaptively. Adaptive queries don't seem to help against our oracle, but rigorously proving this fact seems tricky.

Finally, the lower bounds we prove on the loss of circuit size assume that the distinguishing advantage remains roughly the same. There exist results which are not of this form, in particular – as shown in Figure 2 – the HILL to Metric transformation from [8] only loses in distinguishing advantage, not in circuit size (i.e., we have $s \approx s'$). Proving lower bounds and giving constructions for different circuit size vs. distinguishing advantage trade-offs leave many challenges for future work.

References

1. Theodore Baker, John Gill, and Robert Solovay. Relativizations of the p=?np question. *SIAM Journal on computing*, 4(4):431–442, 1975.
2. Boaz Barak. How to go beyond the black-box simulation barrier. In *42nd FOCS*, pages 106–115. IEEE Computer Society Press, October 2001.
3. Boaz Barak, Ronen Shaltiel, and Avi Wigderson. Computational analogues of entropy. In *In 11th International Conference on Random Structures and Algorithms*, pages 200–215, 2003.
4. Kai-Min Chung, Yael Tauman Kalai, Feng-Hao Liu, and Ran Raz. Memory delegation. In Phillip Rogaway, editor, *CRYPTO 2011*, volume 6841 of *LNCS*, pages 151–168. Springer, August 2011.
5. Yevgeniy Dodis, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In Christian Cachin and Jan Camenisch, editors, *EUROCRYPT 2004*, volume 3027 of *LNCS*, pages 523–540. Springer, May 2004.
6. Stefan Dziembowski and Krzysztof Pietrzak. Leakage-resilient cryptography. In *49th FOCS*, pages 293–302. IEEE Computer Society Press, October 2008.
7. Benjamin Fuller, Adam O'Neill, and Leonid Reyzin. A unified approach to deterministic encryption: New constructions and a connection to computational entropy. In Ronald Cramer, editor, *TCC 2012*, volume 7194 of *LNCS*, pages 582–599. Springer, March 2012.

8. Benjamin Fuller and Leonid Reyzin. Computational entropy and information leakage. Cryptology ePrint Archive, Report 2012/466, 2012. http://eprint.iacr.org/.
9. Craig Gentry and Daniel Wichs. Separating succinct non-interactive arguments from all falsifiable assumptions. In Lance Fortnow and Salil P. Vadhan, editors, *43rd ACM STOC*, pages 99–108. ACM Press, June 2011.
10. Oded Goldreich, Hugo Krawczyk, and Michael Luby. On the existence of pseudorandom generators. *SIAM J. Comput.*, 22(6):1163–1175, 1993.
11. Iftach Haitner, Omer Reingold, and Salil P. Vadhan. Efficiency improvements in constructing pseudorandom generators from one-way functions. In Leonard J. Schulman, editor, *42nd ACM STOC*, pages 437–446. ACM Press, June 2010.
12. Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.
13. Thomas Holenstein. Pseudorandom generators from one-way functions: A simple construction for any hardness. In Shai Halevi and Tal Rabin, editors, *TCC 2006*, volume 3876 of *LNCS*, pages 443–461. Springer, March 2006.
14. Thomas Holenstein and Makrand Sinha. Constructing a pseudorandom generator requires an almost linear number of calls. In *53rd FOCS*, pages 698–707. IEEE Computer Society Press, October 2012.
15. Chun-Yuan Hsiao, Chi-Jen Lu, and Leonid Reyzin. Conditional computational entropy, or toward separating pseudoentropy from compressibility. In Moni Naor, editor, *EUROCRYPT 2007*, volume 4515 of *LNCS*, pages 169–186. Springer, May 2007.
16. Russell Impagliazzo and Steven Rudich. Limits on the provable consequences of one-way permutations. In Shafi Goldwasser, editor, *CRYPTO'88*, volume 403 of *LNCS*, pages 8–26. Springer, August 1988.
17. Dimitar Jetchev and Krzysztof Pietrzak. How to fake auxiliary input. In Yehuda Lindell, editor, *TCC 2014*, volume 8349 of *LNCS*, pages 566–590. Springer, February 2014.
18. Stephan Krenn, Krzysztof Pietrzak, and Akshay Wadia. A counterexample to the chain rule for conditional hill entropy, and what deniable encryption has to do with it. In *10th Theory of Cryptography Conference*, volume 7785, page 23, 2013.
19. Chi-Jen Lu, Shi-Chun Tsai, and Hsin-Lung Wu. On the complexity of hard-core set constructions. In Lars Arge, Christian Cachin, Tomasz Jurdzinski, and Andrzej Tarlecki, editors, *ICALP 2007*, volume 4596 of *LNCS*, pages 183–194. Springer, July 2007.
20. Krzysztof Pietrzak. A leakage-resilient mode of operation. In Antoine Joux, editor, *EUROCRYPT 2009*, volume 5479 of *LNCS*, pages 462–482. Springer, April 2009.
21. Krzysztof Pietrzak and Maciej Skorski. The chain rule for HILL pseudoentropy, revisited. In *Progress in Cryptology - LATINCRYPT 2015 - 4th International Conference on Cryptology and Information Security in Latin America, Guadalajara, Mexico, August 23-26, 2015, Proceedings*, pages 81–98, 2015.
22. Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil P. Vadhan. Dense subsets of pseudorandom sets. In *49th FOCS*, pages 76–85. IEEE Computer Society Press, October 2008.
23. Leonid Reyzin. Some notions of entropy for cryptography - (invited talk). In Serge Fehr, editor, *ICITS 11*, volume 6673 of *LNCS*, pages 138–142. Springer, May 2011.
24. Daniel R. Simon. Finding collisions on a one-way street: Can secure hash functions be based on general assumptions? In Kaisa Nyberg, editor, *EUROCRYPT'98*, volume 1403 of *LNCS*, pages 334–345. Springer, May / June 1998.

25. Maciej Skorski. Metric pseudoentropy: Characterizations, transformations and applications. In Anja Lehmann and Stefan Wolf, editors, *Information Theoretic Security - 8th International Conference, ICITS 2015, Lugano, Switzerland, May 2-5, 2015. Proceedings*, volume 9063 of *Lecture Notes in Computer Science*, pages 105–122. Springer, 2015.
26. Maciej Skorski. A better chain rule for hill pseudoentropy - beyond bounded leakage. In *Information Theoretic Security - 9th International Conference, ICITS 2016*, 2016.
27. Maciej Skorski. Simulating auxiliary information, revisited. In *TCC 2016-B*, 2016.
28. Salil P. Vadhan and Colin Jia Zheng. Characterizing pseudoentropy and simplifying pseudorandom generator constructions. In Howard J. Karloff and Toniann Pitassi, editors, *44th ACM STOC*, pages 817–836. ACM Press, May 2012.
29. Salil P. Vadhan and Colin Jia Zheng. A uniform min-max theorem with applications in cryptography. In Ran Canetti and Juan A. Garay, editors, *CRYPTO 2013, Part I*, volume 8042 of *LNCS*, pages 93–110. Springer, August 2013.
30. Thomas Watson. Advice lower bounds for the dense model theorem. *TOCT*, 7(1):1, 2014.
31. Jiapeng Zhang. On the query complexity for showing dense model. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:38, 2011.

A Proof of Theorem 1

A.1 Majority is best

We prove two statements which are quantitative forms of (i) and (ii) discussed after the statement of Theorem 1. First we show that the probabilistic adversary A^* easily distinguishes X from all Y of high min-entropy.

Claim 1 (Probabilistic Metric Entropy of X is small) *Let A^* be a probabilistic adversary who on input x samples a random $i \in [1,\ldots,h]$, then queries for $D_i(x)$ and outputs the response. Then for any $\delta \leqslant (p-q)/3$ we have*

$$\Pr[\forall Y:\ \mathbf{H}_{\infty}(Y) \geqslant m+1,\ \Delta^{\mathsf{A}^*}(X;Y) \geqslant (p-q)/3] \geqslant 1-2^{\max(n-1,m+1)}\exp(-h\delta^2). \tag{6}$$

Remark 3 (The complexity of the probabilistic distinguisher). We can chose h in Claim 1 to be 2^n, then A^* is of size $O(n)$ and makes only one query.

Consider now a deterministic adversary A who on input x can make at most t queries learning $D_i(x)$ for t different $i \in [1,\ldots,h]$. We claim that

Claim 2 (Deterministic Metric Entropy is big) *Suppose that we have $n \geqslant k+\log(1/\epsilon)+4$ and $\delta = \frac{\epsilon^2}{2+2\epsilon}$. Then for every nonadaptive adversary A which makes $t \leqslant \frac{\ln(2/\epsilon)}{6(p-q)^2}$ queries we have*

$$\Pr_{X,D_1,\ldots,D_h}\left[\exists Y:\ \mathbf{H}_{\infty}(Y) \geqslant k,\ \Delta^{\mathsf{A}}(X;Y) \leqslant \epsilon\right] \geqslant 1-4\exp(-2^m\delta^2). \tag{7}$$

Setting $p-q = 6\epsilon$ we see that Equation (2) follows from Claim 1 and Equation (3) follows from Equation (7) combined with the union bound over all distinguishers. Note that the right hand side of Equation (7) converges to 1 with the rate *doubly* exponential in m, so we can even afford taking a union bound over all distinguishers of size exponential in m.

Proof (of Claim 1). By a Chernoff bound[6] and the union bound

$$\Pr_{X,D_1,\ldots,D_h}[\forall x \in S^c \ : \ \Pr[\mathsf{A}^*(x) = 1] \leqslant q + \delta] \geqslant 1 - 2^{n-1}\exp(-2\delta^2 h) \tag{8}$$

similarly

$$\Pr_{X,D_1,\ldots,D_h}[\forall x \in S : |\Pr[\mathsf{A}^*(x) = 1] - p| \leqslant \delta] \geqslant 1 - 2^m \cdot 2\exp(-2\delta^2 h). \tag{9}$$

The advantage of A^*, with probability $1 - 2^{n-1}\exp(-2h\delta^2)$, equals

$$\begin{aligned} \Delta^{\mathsf{A}^*}(X;Y) &\geqslant (p-\delta) - (p+\delta)\Pr[Y \in S] - (q+\delta)\Pr[Y \in S^c] \\ &\geqslant p - q - (p-q)\Pr[Y \in S] - 2\delta. \end{aligned}$$

Since by the assumption we have $\Pr[Y \in S] \leqslant \frac{1}{2}$, Equation (6) follows.

Proof (of Claim 2). The adversary A non-adaptively queries for $D_i(x)$ values for t distinct i's and then outputs a bit, this bit is thus computed by a function of the form

$$f\left(x, D_{i_1(x)}(x), \ldots, D_{i_t(x)}(x)\right), \tag{10}$$

for some fixed boolean function $f : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}$. We start by simplifying the event (7) using the following proposition, which gives an alternative characterization of the deterministic metric entropy.

Lemma 1 ([3,25]). *Let D be a boolean deterministic function on $\{0,1\}^n$. Then there exists Y of min-entropy at least k such that $\Delta^D(X;Y) \leqslant \epsilon$ if and only if*

$$\mathbb{E}\, D'(X) \leqslant 2^{n-k}\,\mathbb{E}\, D'(U) + \epsilon \tag{11}$$

holds for $D' \in \{D, \mathbf{1} - D\}$

Since $|S^c| \geqslant 2^{n-1}$, we have $\mathbb{E}\, D(U) \geqslant \mathbb{E}_{x \leftarrow S^c} D(x)/2$ for any function D. Therefore, by Lemma 1, the inequality (7) will be proved if we show that the following inequality holds:

$$\Pr_{X,D_1,\ldots,D_h}\left[\forall \mathsf{A}' \in \{\mathsf{A}, \mathbf{1} - \mathsf{A}\} : \mathop{\mathbb{E}}_{x \leftarrow S} \mathsf{A}'(x) \leqslant 2^{n-k-1} \mathop{\mathbb{E}}_{x \leftarrow S^c} \mathsf{A}'(x) + \epsilon\right] \geqslant 1 - 4\exp(-2^m\delta^2) \tag{12}$$

[6] We use the following version: let X_i for $i = 1, \ldots, N$ be independent random variables such that $X_i \in [a_i, b_i]$. Then for any positive t we have $\Pr_{X_1,\ldots,X_N}\left[\sum_{i=1}^N X_i - \mathbb{E}\left[\sum_{i=1}^N X_i\right] \geqslant t\right] \leqslant \exp\left(\frac{2t^2}{\sum_{i=1}^N (b_i - a_i)^2}\right)$.

By the union bound, it is enough to show that for $\mathsf{A}' \in \{\mathsf{A}, \mathbf{1} - \mathsf{A}\}$ we have

$$\Pr_{X, D_1, \ldots, D_h}\left[\mathop{\mathbb{E}}_{x \leftarrow S} \mathsf{A}'(x) \leqslant 2^{n-k-1} \mathop{\mathbb{E}}_{x \leftarrow S^c} \mathsf{A}'(x) + \epsilon\right] \geqslant 1 - 2\exp(-2^m \delta^2) \tag{13}$$

In the next step we simplify the expressions $\mathbb{E}_{x \leftarrow S} \mathsf{A}'(x)$ and $\mathbb{E}_{x \leftarrow S^c} \mathsf{A}'(x)$. The following fact is a direct consequence of the Chernoff bound.

Proposition 1. *For any function $f \in \{0,1\}^n \times \{0,1\}^t \to [0,1]$ we have*

$$\left|\mathop{\mathbb{E}}_{x \leftarrow S} f\left(x, D_{i_1(x)}(x), \ldots, D_{i_t(x)}(x)\right) - \mathbb{E} f(U_n, B_p^1, \ldots, B_p^t)\right| \leqslant \delta \tag{14}$$

$$\left|\mathop{\mathbb{E}}_{x \leftarrow S^c} f\left(x, D_{i_1(x)}(x), \ldots, D_{i_t(x)}(x)\right) - \mathbb{E} f(U_n, B_q^1, \ldots, B_q^t)\right| \leqslant \delta \tag{15}$$

with probability $1 - 2\exp(-2 \cdot 2^m \delta^2)$ over the choice of X and $D_1, \ldots, D_h$.

For any $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_t) \in [0,1]^t$, and any (deterministic or randomized) function $f \in \{0,1\}^t \to [0,1]$ we denote $\mathbb{E}_{\mathbf{r}} f = \mathbb{E} f(B_{\mathbf{r}_1}, \ldots, B_{\mathbf{r}_t})$. It is enough to show that if $\mathbf{r}, \mathbf{r}'$ are both chosen from $\{p, q\}^t$ then we have

$$\mathbb{E}_{\mathbf{r}} f + \delta \leqslant 2^{n-k-1} \max(\mathbb{E}_{\mathbf{r}'} f - \delta, 0) + \epsilon. \tag{16}$$

This inequality will follow by the following lemma (applied to f in the proposition but considered as a function of $\{0,1\}^t$ randomized with the first n input bits).

Lemma 2. *Suppose that $p, q > 0$ are such that $p + q = 1$. Let $f : \{0,1\}^t \to [0,1]$ be an arbitrary function and let $\mathbf{r}, \mathbf{r}' \in \{p, q\}^t$. Then for any $c > 0$ we have*

$$\mathbf{E}_{\mathbf{r}} f \leqslant \exp\left(\frac{(c+1)(p-q)^2}{q} \cdot t\right) \cdot \mathbf{E}_{\mathbf{r}'} f + \exp(-2c^2(p-q)^2 t).$$

Proof. The idea of the proof is to show that for most values of z the ratio $\Pr[B_{\mathbf{r}} = z] / \Pr[B_{\mathbf{r}'} = z]$ is bounded. We have

$$\begin{aligned} &\Pr[B_{\mathbf{r}} = z]/\Pr[B_{\mathbf{r}'} = z] && (17)\\ &= (p/q)^{\#\{i: z_i = 1,\ \mathbf{r}_i > \mathbf{r}'_i\} - \#\{i: z_i = 1,\ \mathbf{r}_i < \mathbf{r}'_i\}} \cdot (q/p)^{\#\{i: z_i = 0,\ \mathbf{r}_i > \mathbf{r}'_i\} - \#\{i: z_i = 0,\ \mathbf{r}_i < \mathbf{r}'_i\}} \\ &= (p/q)^{\#\{i: z_i = 1,\ \mathbf{r}_i > \mathbf{r}'_i\} - \#\{i: z_i = 0,\ \mathbf{r}_i > \mathbf{r}'_i\} - \#\{i: z_i = 1,\ \mathbf{r}_i < \mathbf{r}'_i\} + \#\{i: z_i = 0,\ \mathbf{r}_i < \mathbf{r}'_i\}} \\ &= (p/q)^{\sum_{i=1}^{t} (2z_i - 1) \cdot \mathrm{sgn}(\mathbf{r}_i - \mathbf{r}'_i)} && (18) \end{aligned}$$

The random variables $\xi_i = (2z_i - 1) \cdot \mathrm{sgn}(\mathbf{r}_i - \mathbf{r}'_i)$ for $i = 1, \ldots, t$, where z is sampled from $B_{\mathbf{r}}$, are independent with the expectations $\mathbb{E}\xi_i = (2\mathbf{r}_i - 1)\mathrm{sgn}(\mathbf{r}_i - \mathbf{r}'_i) \leqslant p - q$. By the Chernoff bound for any $c > 0$ we get

$$\Pr_{z \leftarrow B_{\mathbf{r}}}\left[\sum_{i=1}^{t} (2z_i - 1) \cdot \mathrm{sgn}(\mathbf{r}_i - \mathbf{r}'_i) \geqslant (p-q)t + c(p-q)t\right] \leqslant \exp(-2c^2(p-q)^2 t). \tag{19}$$

Therefore,

$$\mathbb{E}_{\mathbf{r}} f \leqslant (p/q)^{(c+1)(p-q)t} \mathbb{E}_{\mathbf{r}'} f + 2\exp(-2c^2(p-q)^2 t) \tag{20}$$

and the claim follows by observing that $p/q = 1 + (p-q)/q \leqslant \exp((p-q)/q)$.

From Lemma 2 it follows that Equation (16) is satisfied with

$$\delta \leqslant \frac{\epsilon}{2\exp\left((c+1)(p-q)^2 \cdot t/q\right)+2} \tag{21}$$

provided that

$$\exp(-2c^2(p-q)^2 \cdot t) \leqslant \epsilon/2 \tag{22}$$

$$\exp\left((c+1)(p-q)^2 \cdot t/q\right) \leqslant 2^{n-k-1} \tag{23}$$

It is easy to see that Equation (23) and Equation (22) are satisfied if and only if

$$\frac{\ln(2/\epsilon)}{2c^2(p-q)^2} \leqslant t \leqslant (n-k-3)\ln 2 \cdot \frac{q}{(c+1)(p-q)^2}.$$

This inequality can be satisfied if and only if

$$\epsilon \geqslant 2 \cdot 2^{(k-n+3)\cdot\frac{2qc^2}{c+1}}.$$

If we set $t = \frac{\ln(2/\epsilon)}{2c^2(p-q)^2}$ then Equation (21) becomes

$$\delta \leqslant \frac{\epsilon}{(2/\epsilon)^{\frac{c+1}{2qc^2}}+2}$$

Choosing c so that $\frac{2qc^2}{c+1} = 1$ we see that it is enough to assume $\epsilon \geqslant 2 \cdot 2^{k-n+3}$, any δ such that $\delta \leqslant \frac{\epsilon^2}{2+2\epsilon}$ and $t \approx \frac{\ln(2/\epsilon)}{6(p-q)^2}$ (the constant 6 is sightly bigger than the exact value, but if Claim 2 holds true for some t then also for $t' < t$). This finishes the proof of Claim 2.

B Proof of Theorem 2

A Remark on The Oracle. For convenience, the oracle $\mathcal{O} : \{0,1\}^n \rightarrow \{0,1\}$ we use in the proof is probabilistic, in the sense that it flips some random coins before answering a query (in particular, making the same query twice might give different outputs). We remark that, as the adversaries considered are probabilistic, one can replace this oracle with a deterministic one $\mathcal{O}_{\mathrm{det}}$ by assigning to every possible query x a 2^L tuple $(x,r), r \in \{0,1\}^L$ of queries (for some sufficiently large L), where the output for $\mathcal{O}_{\mathrm{det}}((x,r))$ is sampled according to $\mathcal{O}(x)$ for every r. We can emulate the output distribution $\mathcal{O}(x)$ by querying $\mathcal{O}((x,r))$ for a random r. On the other hand, for a random x, even an exponential size distinguisher will not be able to distinguish $\mathcal{O}_{\mathrm{def}}((x,\cdot))$ from an oracle which, when queried on input (x,r) for the first time, samples the output according to the distribution of $\mathcal{O}(x)$.[7]

[7] This can be shown along the lines of the proof that a random exponential size subset is unconditionally pseudorandom against exponential size distinguishers, see Goldreich's book "Foundations of Cryptography – Basic Techniques", Proposition 3.2.3.

Proof (of Theorem 2). We first describe how we construct the distribution (X, Z) and the oracle $\mathcal{O}$.

Construction details. We chose at random two disjoint sets $\mathcal{X}_0, \mathcal{X}_1 \subset \{0,1\}^n$ of size 2^ℓ and define $\mathcal{X} = \mathcal{X}_0 \cup \mathcal{X}_1$. Let $\pi : \{0,1\}^n \to \{0,1\}^{m-1}$ be a random function. The oracle $\mathcal{O}$ on input $(x, j) \in \mathcal{X} \times \{0,1\}^{m-1}$ outputs a sample of $B_{1/2}$ (i.e., a uniformly random bit), except if $x \in \mathcal{X}$ and $\pi(x) = j$, in this case the output bit has bias δ; If $x \in \mathcal{X}_0$, the oracle outputs a sample of $B_{1/2-\delta}$, and otherwise, if $x \in \mathcal{X}_1$, a sample of $B_{1/2+\delta}$. We define the joint distribution (X, Z) by sampling $Z' \leftarrow \{0,1\}, X \leftarrow \mathcal{X}_{Z'}$ and setting $Z = \pi(X) \| Z'$ (note that X is uniform in $\mathcal{X}$)

Adversaries. The adversary on input $x \in \{0,1\}^n$ makes T non-adaptive queries $(x, j_1(x)), \ldots, (x, j_T(x))$ to the oracle. We denote $\mathcal{O}$'s response with $R(x) = \left(R^i(x, j_i(x))\right)_{i=1}^{T}$. The adversary's final output $f(x, R(x))$ is computed by a boolean function $f : \{0,1\}^n \times \{0,1\}^T \to \{0,1\}$.

Formal proof. Let $R(x) = (R^1(x, j_1(x)), \ldots, R^T(x, j_T(x)))$ be the sequences of the oracle's responses and Let $B(x) = (B^1_{1/2}, \ldots, B^T_{1/2})$ be independent random bits. For every x the number of *useful* responses, that is indexes i such that $R^i(x, j_i(x))$ is biased, is defined to be

$$T(x) = \sum_{i=1}^{T} [j_i(x) = \pi(x)] \tag{24}$$

On average we have $\mathbb{E}_{\mathcal{O}(\cdot)} T(x) = T/2^{m-1}$. We claim that the adversary actually learns basically nothing about $\mathcal{X}$: the sequence of oracle outptus is close to the sequence of unbiased bits. We start by showing that $\mathcal{X}$ is pseudorandom for our adversary.

Claim 3 (X is pseudorandom, even given oracle responses) *For any f and $\epsilon > 0$ we have*

$$\left| \mathop{\mathbb{E}}_{x \leftarrow \mathcal{X}} f(x, R(x)) - \mathop{\mathbb{E}}_{x \leftarrow U_n} f(x, R(x)) \right| \leq \epsilon + O\left(\delta^2 T / 2^m\right) \tag{25}$$

with error probability at most $O\left(\exp\left(-\Omega\left(2^{n-m}\right)\right) + \exp\left(-\Omega\left(2^\ell \epsilon^2\right)\right)\right)$.

Proof. By Lemma 3 and the definition of $\mathcal{O}$, for every $x \in \mathcal{X}$ we obtain

$$|\mathbb{E} f(x, R(x)) - \mathbb{E} f(x, B(x))| = \begin{cases} O\left(T(x)\delta^2\right), & x \in \mathcal{X} \\ 0, & x \notin \mathcal{X} \end{cases} \tag{26}$$

for every boolean function f and some absolute constant hidden under big-Oh. Thus

$$\left| \mathop{\mathbb{E}}_{x \leftarrow \mathcal{X}} f(x, R(x)) - \mathop{\mathbb{E}}_{x \leftarrow \mathcal{X}} f(x, B(x)) \right| = O\left(\mathop{\mathbb{E}}_{x \leftarrow \mathcal{X}} T(x)\delta^2 \right) \tag{27}$$

Note that the random variables $f(x, R(x))$ for different values of x are independent and similarly $f(x, B(x))$ for different values of x are independent. Since the set $\mathcal{X}$ is chosen at random by the Hoeffding-Chernoff bound we obtain that with probability $1 - 2\exp\left(-\Omega\left(2^{\ell}\epsilon^2\right)\right)$ over $\mathcal{O}$ the following holds:

$$\left| \mathop{\mathbb{E}}_{x \leftarrow \mathcal{X}} f(x, B(x)) - \mathop{\mathbb{E}}_{x \leftarrow U_n} f(x, B(x)) \right| \leqslant \epsilon \tag{28}$$

Combining Equation (27) and Equation (28) we obtain (with probability $1 - 2\exp\left(-\Omega\left(2^{\ell}\epsilon^2\right)\right)$ over $\mathcal{O}$)

$$. \left| \mathop{\mathbb{E}}_{x \leftarrow \mathcal{X}} f(x, R(x)) - \mathop{\mathbb{E}}_{x \leftarrow U_n} f(x, B(x)) \right| \leqslant \epsilon + O\left(\mathop{\mathbb{E}}_{x \leftarrow \mathcal{X}} T(x)\delta^2 \right) \tag{29}$$

By Equation (26) we have

$$\left| \mathop{\mathbb{E}}_{x \leftarrow U_n} f(x, R(x)) - \mathop{\mathbb{E}}_{x \leftarrow U_n} f(x, B(x)) \right| \leqslant O\left(\mathop{\mathbb{E}}_{x \leftarrow U_n} T(x)\delta^2 \right). \tag{30}$$

Now Equations (29) and (30) imply

$$\left| \mathop{\mathbb{E}}_{x \leftarrow \mathcal{X}} f(x, R(x)) - \mathop{\mathbb{E}}_{x \leftarrow U_n} f(x, R(x)) \right| \leqslant \epsilon + O\left(\mathop{\mathbb{E}}_{x \leftarrow U_n} T(x)\delta^2 \right). \tag{31}$$

The random variables $T(x)$ for different x are independent, bounded by T and have the first moment $\mathbb{E}_{\mathcal{O}}(T(x)) = T/2^{m-1}$. By the multiplicative Chernoff bound with probability $1-2\exp\left(-\Omega\left(2^{n-m}\right)\right)$ over $\mathcal{O}$ it holds that $\mathbb{E}_{x \leftarrow U_n} T(x) < 2 \cdot T/2^{m-1}$. This implies Equation (25) with error probability at most

$$P_{\text{err}} = O\left(\exp\left(-\Omega\left(2^{n-m}\right)\right) + \exp\left(-\Omega\left(2^{\ell}\epsilon^2\right)\right)\right).$$

Claim 4 *There exists a distinguisher* $\mathsf{D} : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}$ *which calls the oracle* $\mathcal{O}$ *one time and such that for any joint distribution* Y, Z' *over* $\{0,1\}^n \times \{0,1\}^m$ *with entropy* $\widetilde{\mathbf{H}}_{\infty}(Y|Z') \geqslant \ell + 1$ *it holds that*

$$\mathbb{E}\,\mathsf{D}(X, Z) - \mathbb{E}\,\mathsf{D}(Y, Z') \geqslant \frac{\delta}{2}$$

with probability $1 - 2\exp(-\Omega\left(2^{\ell}\delta^2\right))$.

Remark 4 (Amplified distinguisher). Assuming that T is sufficiently large, we can modify D by taking the majority vote over T queries on $\mathcal{O}(x, z)$. This will boost the distinguishing advantage from $\delta/2$ to $C\delta$ where C can be an arbitrary constant (for sufficiently small δ).

Proof (of Claim 4). The distinguisher D simply calls the oracle $\mathcal{O}$ on the pair (x, z). The probability that D outputs 1 on input (Y, Z') is at most (the probabilities below are over the choice of $\mathcal{O}$ and Y, Z')

$$\begin{aligned}
\Pr\left(\mathsf{D}(Y,Z')=1\right) &= \mathop{\mathbb{E}}_{z\leftarrow Z'} \Pr\left(\mathsf{D}(Y|_{Z'=z},z)=1\right) \\
&= \mathop{\mathbb{E}}_{z\leftarrow Z'} \left[\Pr\left(\, \mathsf{D}(Y,z)=1 \wedge Y \notin \mathcal{X} |\, Z'=z\right)\right] + \\
&\quad + \mathop{\mathbb{E}}_{z\leftarrow Z'} \left[\Pr\left(\, \mathsf{D}(Y,z)=1 \wedge Y \in \mathcal{X} |\, Z'=z\right)\right] \\
&= \frac{1}{2} + \delta \cdot \mathop{\mathbb{E}}_{z\leftarrow Z'} \left[\Pr\left(Y \in \mathcal{X} |\, Z'=z\right)\right] \\
&\leqslant \frac{1}{2} + \delta \mathop{\mathbb{E}}_{z\leftarrow Z'} \left[|\mathcal{X}| \cdot 2^{-\mathbf{H}_\infty\left(Y|Z'=z\right)}\right] \\
&= \frac{1}{2} + \delta \cdot |\mathcal{X}| \cdot 2^{-\widetilde{\mathbf{H}}_\infty(Y|Z')}
\end{aligned}$$

which is at most $\frac{1}{2}+\frac{\delta}{2}$. On the other hand we have $\Pr(\mathsf{D}(X,Z)=1) = \frac{1}{2}+\delta$. From this we see that the advantage is δ on average - but we need stronger concentration guarantees. Note that $\Pr(\mathsf{D}(X,Z)=1) = \sum_{x\in S}\Pr[X=x]\cdot \mathsf{D}(x,i(x))$ can be viewed as a sum of independent random variables. By the Chernoff-Hoeffding bound we get

$$\Pr_{\mathcal{O}}\left[\Pr(\mathsf{D}(X,Z)=1) \geqslant \frac{1}{2}+\delta-\frac{\delta}{8}\right] \geqslant 1-\exp(-\Omega\left(2^{\ell}\delta^2\right)))$$

Similarly, $\Pr(\mathsf{D}(Y,Z')=1) = \sum_{x,z}\Pr[Y=x,Z'=z]\cdot \mathsf{D}(x,z')$. Since

$$\begin{aligned}
\sum_{x,z}\Pr[Y=x,Z'=z]^2 &= \sum_{z}\sum x \Pr[Z'=z]^2 \Pr[Y=x|Z'=z]^2 \\
&\leqslant \sum_{z}\Pr[Z'=z]2^{-\mathbf{H}_\infty(Y|_{Z'=z})} \\
&\leqslant 2^{-\widetilde{\mathbf{H}}_\infty(Y|Z)},
\end{aligned}$$

the Chernoff-Hoeffding bound implies

$$\Pr_{\mathcal{O}}\left[\Pr(\mathsf{D}(Y',Z)=1) \leqslant \frac{1}{2}+\frac{\delta}{2}+\frac{\delta}{8}\right] \geqslant 1-\exp(-\Omega\left(2^{\ell}\delta^2\right)) \tag{32}$$

and the result follows.

We set $\epsilon = \frac{\delta}{3}$ and $T = c\cdot 2^m/\epsilon$. Now Claim 4 directly implies Equation (5) whereas Equation (4) follows, when c is sufficiently small, from Claim 3 by a union bound; To see this, note that the right hand side of (32) is doubly exponentially close (in ℓ) to 1, and recall that $\ell > m + 6\log(1/\delta)$. So we can take a union bound over all $O(\exp(T))$circuits D of size T and deduce that with high probability the left hand side of (32) hold for all of them.

C Proof of Lemma 3

Lemma 3 (Lower bounds on the coin problem). *Fix $\delta \in (0, 1/2)$ and define $p = \frac{1}{2}+\delta$ and $q = 1-p$. Consider the following two experiments:*

(a) We flip a fair coin, and depending on the result we toss T times a biased coin B_p (probability of the head is p) or toss T times a coin B_q (probability of the head is q). The output is the result of these T flips.

(b) We flip T times a fair coin and output the results.

Then one cannot distinguish (a) from (b) better than with advantage $O\left(T\delta^2\right)$.

Remark 5. We give a simple proof based on calculating Renyi divergences. This result can be also derived by more sophisticaed techniques from Fourier analysis (the generalized XOR lemma).

Before we give the proof, let's recall some basic facts about *Pearson Chi-Squared Distance*. For any two distributions P, Q over the same space, their Chi-Squared distance defined by

$$D_{\chi^2}(P \parallel Q) = \sum_x Q(x)\left(\frac{P(x)}{Q(x)} - 1\right)^2 = \sum_x \frac{P(x)^2}{Q(x)^2} - 1 \tag{33}$$

Now let $U_1, \ldots, U_n$ be independent uniform bits, $X_1, \ldots, X_n$ be i.i.d. bits where 1 appears with probability $p = \frac{1}{2} + \delta$ and $Y_1, \ldots, Y_n$ be i.i.d. bits where 1 appears w ith probability $q = 1 - p = \frac{1}{2} - \delta$. We want to estimate the distance between $U = U_1, \ldots, U_n$ and Z distributed as an equally weighted combination of $X = X_1, \ldots, X_n$ and $Y = Y_1, \ldots, Y_n$. We think of δ as a fixed parameter and n as a growing number. Our statement will easily follow by combining the following two claims

Claim 5 *With U and Z as above, and for $n = O\left(\delta^{-2}\right)$, it holds that*

$$D_{\chi^2}\left(U; Z\right) = O\left(n^2\delta^4\right) \tag{34}$$

Claim 6 *For any R and uniform U*

$$\mathrm{SD}(R \parallel U) \leqslant \sqrt{D_{\chi^2}(R \parallel U)}, \tag{35}$$

Indeed, combining these claims we obtain $\mathrm{SD}(Z \parallel U) = O(n\delta^2)$ when $n = O\left(\delta^{-2}\right)$. Since the left-hand side is bounded by 1, this is true also when $n > c\delta^{-2}$ for some absolute constant c and the result follows.

Proof (of Claim 5). We have

$$
\begin{aligned}
D_{\chi^2}\left(\frac{1}{2}P_{X_1,\ldots,X_n} + \frac{1}{2}P_{Y_1,\ldots,Y_n} \parallel P_{U_1}\cdot\ldots\cdot P_{U_n}\right) = \\
2^n \cdot \sum_{z_1,\ldots,z_n}\left(\frac{1}{2}P_{X_1}(z_1)\cdot\ldots\cdot P_{X_n}(z_n) + \frac{1}{2}P_{Y_1}(z_1)\cdot\ldots\cdot P_{Y_n}(z_n)\right)^2 - 1 = \\
\frac{1}{4}\cdot 2^n \prod_i \left(\sum_z P_{X_i}(z)^2\right) + \frac{1}{4}\cdot 2\cdot 2^n \prod_i \left(\sum_z P_{X_i}(z)P_{Y_i}(z)\right) + \\
+\frac{1}{4}\cdot 2^n \prod_i \left(\sum_z P_{Y_i}(z)^2\right) - 1 = \\
\frac{1}{4}\left((1+4\delta^2)^n + 2(1-4\delta^2)^n + (1+4\delta^2)^n - 4\right)
\end{aligned}
\tag{36}
$$

and the result follows by the Taylor expansion $(1+u)^n = 1 + nu + O(n^2u^2)$ where $nu = O(1)$ applied to $u = 4\delta^2$. The bound is valid as long as $n = O\left(\delta^{-2}\right)$.

Proof (of Claim 6). This inequality follows immediately from the Cauchy-Schwarz inequality and the definition of D_{χ^2}.