Best non-uniform attacks against pseudoentropy

Krzysztof Pietrzak, Maciej Skorski

IST Austria

Abstract. We prove that psuedoentropy amount k can be broken in advantage ϵ and size $O(2^k\epsilon^2)$. This extends best attacks against pseudorandomness due to De, Trevisan and Tulsiani.

1 Introduction

De, Trevisan and Tulsiani [DTT10] show how to construct a non-uniform attack against any given n-bit seed PRG, which achieves advantage ϵ in time $s = O\left(2^n\epsilon^2\right)$. This work addresses a related question about the existence of fundamental limits for pseudoentropy quality. We prove a strengthening of the former result, showing a non-uniform attack against k bit of pseudoentropy with any advantage ϵ and time $s = O\left(2^k\epsilon^2\right)$. In this setting, the attack means that one can distinguish between a given distribution and every distribution of min-entropy k. This may be interesting because

1. We provide a clear separation between pseudoentropy and information-theoretic smooth min-entropy. In particular we see what are adversarial resources for which *computational gain* in the entropy amount is possible.
2. We rule out the existence of pseudorandomness condensers (which are interesting in view of recent works on key derivation from weak sources) for certain security parameters.

1.1 Our contribution

Theorem 1. *Suppose that X does not have k bits of δ-smooth min-entropy. Then for any ϵ we have*

$$H^{\mathrm{HILL}}_{\tilde{O}(2^k\epsilon^2\delta^{-2}),\Omega(\epsilon)}(X) < k$$

for some universal constant hidden under $\Omega(\cdot)$ and a factor linear in n hidden under $\tilde{O}(\cdot)$.

The following somewhat less general corollary makes this statement more clear

Corollary 1. *Let f be a deterministic function. Then we have*

$$H^{\mathrm{HILL}}_{\tilde{O}(2^k\epsilon^2),\Omega(\epsilon)}(f(U_k)) < k+1.$$

1.2 Proof outline

A weaker result as a ball-bins problem We outline the proof of a somewhat weakened version of Corollary 1 in the language of balls and bins. For every Y of min-entropy $k' = k + O(1)$ we want to distinguish Y from $X = f(U_k)$. Suppose for simplicity that Y is flat and f is injective, so that X is also flat. Our strategy will be to hash the points randomly into two bins and take advantage of the fact that the *average maximum load* is closer to $\frac{1}{2}$ when we sample from Y than when drawing from X. The reason is that Y has more bins so that after averaging the load is somewhat "more concentrated" around the mean.

Think of throwing balls (inputs x) into two bins (labeled by -1 and 1). If the balls come from the support of X, the maximum load (over two bins) equals $2^{k-1} + O\left(2^{k/2}\right)$ with high probability. Similarly, if the balls come from the support of Y, then maximum load is $2^{k'-1} + O\left(2^{k'/2}\right)$ with high probability. In terms of the average load (the load normalized by the total number of balls)

$$\mathsf{AverageMaxLoad}(X) = 0.5 + \Theta\left(2^{-k/2}\right) \qquad \text{w.h.p. when drawing from } X$$
$$\mathsf{AverageMaxLoad}(Y) = 0.5 + \Theta\left(2^{-k'/2}\right) \qquad \text{w.h.p. when drawing from } Y$$

By choosing k' so that $k' - k \gg 0$ we obtain (with positive probability)

$$\mathsf{AverageMaxLoad}(X) - \mathsf{AverageMaxLoad}(Y) = \Omega(2^{-k/2}).$$

Letting D be one of these bins assignments we obtain a distinguisher with advantage $\epsilon = \Omega(2^{-k/2})$. To generate the assignments efficiently we relax the assumption about choosing bins and assume only that the choices of bins are independent for any group of $\ell = 4$ balls. The fourth moment method allows us to keep sufficiently good probabilistic guarantees on the maximum load.

The general case by random walks techniques

A high-level outline and comparison to [DTT10] Below in Figure 1 we sketch the flow of the argument.

At a very high level it is similar to the one given in [DTT10]. The core idea is that a random mapping D of the domain to $\{-1, 1\}$ likely distinguishes any two n-bit distributions X and Y with advantage being the euclidean distance between probability functions of X and Y.

For any X and Y separated in the statistical distance by a constant (this is the setting for PRGs applications) this yields a bound $\Omega\left(2^{-\frac{n}{2}}\right)$. This bound can be then amplified, at the cost of extra advice, by partitioning the domain and combining corresponding advantages (advice basically encodes if there is a need for flipping the output). Finally one can show that 4-wise independence provides enough randomness for this argument, which makes sampling D efficient. Our argument deviates from this approach in two important aspects.

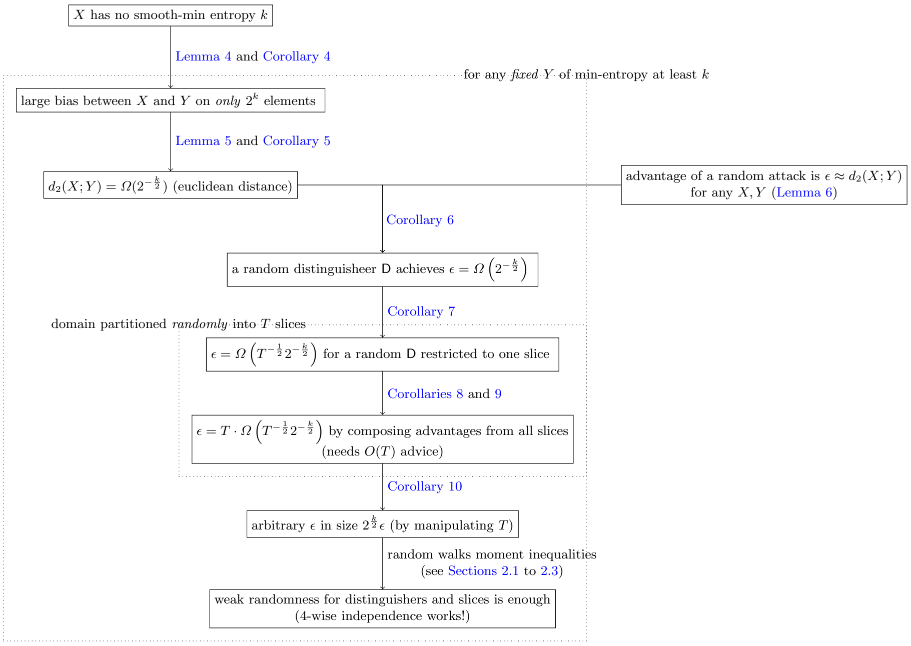

X has no smooth-min entropy k
Lemma 4 and Corollary 4
for any *fixed* Y of min-entropy at least k
large bias between X and Y on *only* 2^k elements
Lemma 5 and Corollary 5
$d_2(X;Y) = \Omega(2^{-\frac{k}{2}})$ (euclidean distance)
advantage of a random attack is $\epsilon \approx d_2(X;Y)$ for any X, Y (Lemma 6)
Corollary 6
a random distinguisheer D achieves $\epsilon = \Omega\left(2^{-\frac{k}{2}}\right)$
Corollary 7
domain partitioned *randomly* into T slices
$\epsilon = \Omega\left(T^{-\frac{1}{2}}2^{-\frac{k}{2}}\right)$ for a random D restricted to one slice
Corollaries 8 and 9
$\epsilon = T \cdot \Omega\left(T^{-\frac{1}{2}}2^{-\frac{k}{2}}\right)$ by composing advantages from all slices (needs $O(T)$ advice)
Corollary 10
arbitrary ϵ in size $2^{\frac{k}{2}}\epsilon$ (by manipulating T)
random walks moment inequalities (see Sections 2.1 to 2.3)
weak randomness for distinguishers and slices is enough (4-wise independence works!)

Fig. 1: The map of our proof.

The first difference is that, as we show, in the pseudoentropy case one can improve the advantage of a random distinguisher to $\Omega\left(2^{-\frac{k}{2}}\right)$. The reason is that being statistically away from k-bit min-entropy distributions implies a *large bias on already* 2^k *elements*. This fact (see Lemma 4 and Corollary 4, and also Figure 3) is a new characterization of smooth min-entropy of independent interest.

The second subtlety arises when it comes to amplify the advantage over the partition slices. For the pseudorandomness case it is enough to split the domain deterministically, for example by fixing prefixes of n-bit strings. In our case this leads to the loss of the improvement $\Omega\left(2^{-\frac{k}{2}}\right)$. For us a "good" partition must shatter our 2^k-element high-biased set. Since this set can be arbitrary complicated (contrarily to the pseudorandomness case!), the solution is to use *random partitions* and simulate them by 4-universal hashing. Generating base distinguishers and partitions at the same time makes probability calculations more involved.

Technical calculations are based on the fourth moment method, similarly as in [DTT10]. The base idea is that for settings where the second and fourth moment are easy to compute (e.g. sums of independent symmetric random variables) we can obtain good upper and lower bounds on the first moment. In the context of algorithmic applications these techniques are usually credited to [Ber97]. Interestingly, exploiting natural relations to *random walks*, we show that calculations immediately follow by adopting classical (almost one century old) tools and results [MZ37, Khi24]. Our technical novelty is an application of moment inequalities due to Marcinkiewicz-Zygmund and Paley-Zygmund, which allow us to prove slightly more than just the existence of an attack. Namely we generate it with constant success probability. While there is no general way to verify the success, we find this statement interesting because of techniques applied, and closer to what happens for D chosen truly uniformly.

Advantage $\Omega(2^{-k/2})$ in linear time. For the sake of contradiction we assume that the δ-smooth min-entropy of X is smaller than k. This assumption can be seen as a statement about the "shape" of the distribution. Namely, one can show that the mass of X that is above the threshold 2^{-k} equals at least δ, that is

$$\sum_x \max(P_X(x) - 2^{-k}, 0) \geqslant \delta.$$

For an illustration see Figure 3.

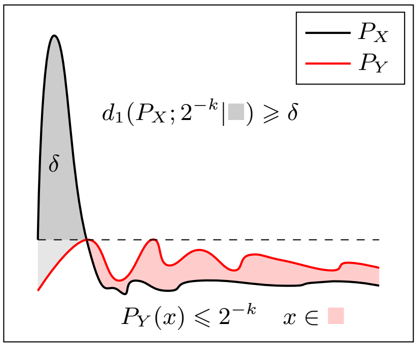

P_X
P_Y
$d_1(P_X; 2^{-k}|▨) \geqslant \delta$
δ
$P_Y(x) \leqslant 2^{-k}$
$x \in$ ▨

Fig. 2: An intuition behind the attack. Random ±1-weights make the bias equal to the ℓ_2-distance of P_X and P_Y. This distance can be bounded in terms of the ℓ_1 distance, which concentrates mass difference δ on less than 2^k elements (the region in shadow gray).

Now we are ready to construct an attack. We define the advantage of a function D on n bits as

$$\mathsf{Adv}^{\mathsf{D}} = \left| \sum_x \mathsf{D}(x)(P_X(x) - P_Y(x)) \right|$$

(writing also $\mathsf{Adv}^{\mathsf{D}}$ when the summation is restricted to a subset S). Consider a distinguisher D which for every x randomly outputs $\{-1, 1\}$. It's easy to see that the second moment of this advantage (under the choice of D) equals

$$\mathbb{E}\left(\mathsf{Adv}^{\mathsf{D}}\right)^2 = \sum_x (P_X(x) - P_Y(x))^2 = d_2(P_X; P_Y).$$

By elementary inequalities and what we know about the bias between P_X and P_Y (namely that it bigger than δ on 2^k elements, see Figure 3), we show that

$$d_2(P_X;P_Y) \geqslant 2^{-\frac{k}{2}}\delta,$$

which combined with the last inequality proves that the advantage is $2^{-\frac{k}{2}}\delta$ for at least one choice of D. Finally, since we used only second moment with calculations involving, it suffices to generate $\mathsf{D}(x)$ as pairwise independent random variables, which costs $O(n)$. By assuming 4-wise independence we can prove slightly more, namely that a constant fraction of generated D's are good distinguishers. While this assumption is not necessary at this step, it will be important to amplify the advantage.

Leveraging the advantage by slicing the domain Consider a random and equitable partition $\{S_i\}_{i=1}^{T}$ of the set $\{0,1\}^n$. From the previous analysis we know that a random distinguisher achieves advantage $\epsilon = d_2(P_X;P_Y)$ over the whole domain. Note that (for any, not necessarily random partition $\{S_i\}_i$) we have

$$(d_2(P_X;P_Y))^2 = \sum_{i=1}^{T} (d_2(P_X;P_Y|S_i))^2$$

where $d_2(P_X;P_Y|S_i)$ is the restriction of the distance to the set S_i (by restricting the summation to S_i). From a random partition we expect the mass difference between P_X and P_Y to be *distributed evenly* among the partition slices (see Figure 3(b)). Based on the last equation, we expect

$$d_2(P_X;P_Y|S_i) \approx \frac{d_2(P_X;P_Y)}{\sqrt{T}}$$

to hold with high probability over $\{S_i\}_i$. In fact, if the mass difference is heavily unbiased on partition slices (for example concentrated on one slice) our argument will not offer any gain over the previous construction (see see Figure 3(a)).

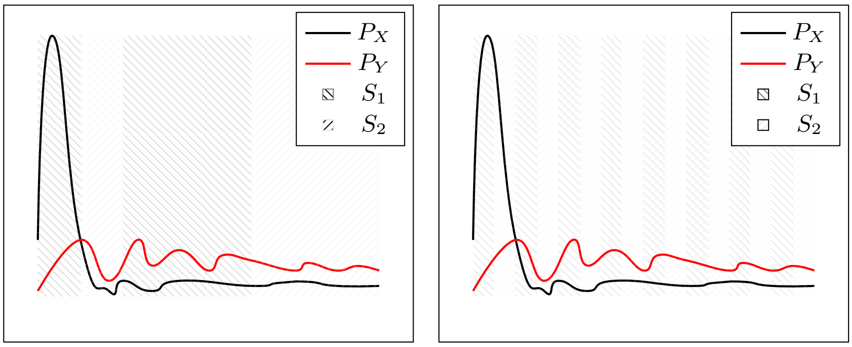
By applying the previous argument to individual slices, for every i we obtain a distinguisher D_i with advantage $\mathsf{Adv}_{S_i}^{\mathsf{D}_i} = \Omega\left((T^{-\frac{1}{2}}2^{-\frac{k}{2}})\delta\right)$ restricted to the set S_i (with high probability over the choice of D and $\{S_i\}_i$). Now if the sets S_i are *efficiently recognizable*, we can combine them into a better distinguisher. Namely if we define

$$\hat{\mathsf{D}}(x) = \mathsf{D}_i(x), \text{ where } i \text{ is such that } x \in S_i,$$

then the advantage equals (with high probability over the random D and $\{S_i\}_i$)

$$\mathsf{Adv}^{\hat{\mathsf{D}}} = \sum_{i=1}^{T} \mathsf{Adv}_{S_i}^{\mathsf{D}_i} = \Omega\left(T^{\frac{1}{2}}2^{-\frac{k}{2}}\delta\right)$$

and the complexity equals $\tilde{O}(T)$ where the hidden constant is the cost of deciding S_i. By changing T we get a smooth tradeoff $s = \tilde{O}(2^k\epsilon^2\delta^{-2})$ between the



(a) An example of a “bad” partition. Almost all advantage in one slice S_1.

(b) An example of a “good“ partition. The advantage is evenly distributed among slices S_1, S_2.

Fig. 3: Bad and good partitions.

advantage ϵ and the circuit size s. This discussion shows that to complete the argument we need the following two properties of the partition π: (a) the mass difference between P_X and P_Y is (roughly) equidistributed among slices and (b) the membership in partition slices can be efficiently decided.

Making slices efficient and enough random To complete the argument, we generate π by seeding a 4-universal hash function from n bits to $t = \lceil \log(T) \rceil$ bits. The i-th slice S_i is defined as the preimage of the i-th output (in the lexicographic order) and can be computed in size $O(n)$. We also require $\mathsf{D}(x)$ to be generated in a 4-wise independent manner. These assumptions are enough to prove that

$$\mathbb{E}\,\mathsf{Adv}^{\mathsf{D}}_{S_i} = \varOmega\left(T^{-\frac{1}{2}} d_2(P_X;P_Y)\right) = \varOmega\left(T^{-\frac{1}{2}} 2^{-\frac{k}{2}}\right).$$

Interestingly, the expected advantage (left-hand side) cannot be computed directly. The trick here is to bound it in terms of the second and fourth moment.

The above inequality, coupled with bounds on second moments of the advantage $\mathsf{Adv}^{\mathsf{D}}_{S_i}$ (obtained directly), allows us to prove that

$$\Pr\left[\sum_{i=1}^{T} \mathsf{Adv}^{\mathsf{D}}_{S_i} \geqslant \varOmega(1) \cdot T^{\frac{1}{2}} 2^{-\frac{k}{2}} \delta\right] > 0.$$

1.3 More about proof techniques

From a technical point of view, our method involves computing higher moments of the advantages to obtain concentration and anti-concentration results. The key observation is that the advantage written down as

$$\mathsf{Adv}^{\mathsf{D}}(S_i) = \left| \sum_{x} (P_X(x) - P_Y(x)) \mathbf{1}_{S_i}(x) \mathsf{D}(x) \right|$$

which can be then studied as a *random walk*

$$\mathsf{Adv}^{\mathsf{D}}(S_i) = \left|\sum_{i,x} \xi_{i,x}\right|$$

with zero-mean increments $\xi_{i,x} = (P_X(x) - P_Y(x))\mathbf{1}_{S_i}(x)\mathsf{D}(x)$. The difference with respect to classical model is that the increments are only ℓ-wise independent (for $\ell = 4$). We observe however, that some of the classical moment bounds still apply to this setting (see Sections 2.2 and 2.3 for more details).

2 Preliminaries

2.1 Interpolation Inequalities

Interpolation inequalities show how to bound the p-th moment of a random variable if we know bounds on one smaller and one higher moment.

Lemma 1 (Moments interpolation). *For any $p_1 < p < p_2$ and any bounded random variable Z we have*

$$\|Z\|_p \leqslant \left(\|Z\|_{p_1}\right)^{\theta}\left(\|Z\|_{p_2}\right)^{1-\theta}$$

where θ is such that $\frac{\theta}{p_1} + \frac{1-\theta}{p_2} = \frac{1}{p}$, and for any r we define $\|Z\|_r = \left(\mathbb{E}\,|Z|^r\right)^{\frac{1}{r}}$.

Alternatively, we can *lower bound* a moment given *two higher moments.* This is very useful when higher moments are easier to compute. In this work will bound from below first moments when we know the second and the fourth moment (easier to compute as they are even-order moments)

Corollary 2. *For any bounded Z we have $\mathbb{E}\,|Z| \geqslant \frac{\left(\mathbb{E}\,|Z|^2\right)^{\frac{3}{2}}}{\left(\mathbb{E}\,|Z|^4\right)^{\frac{1}{2}}}$.*

2.2 Moments of random walks

For a random walk $\sum_x \xi(x)$ where $\xi(x)$ are independent with zero-mean we have good control over the moments, namely $\mathbb{E}\,|\sum_x \xi(x)|^p = \Theta(1)\cdot\left(\sum_x \mathrm{Var}(\xi(x))\right)^{\frac{p}{2}}$ where constants depend on p. This result is due to Marcinkiewicz and Zygmund [MZ37] who extended the less general result of Khintchine [Khi24]. Below we notice that for *small moments p* it suffices to assume only p-wise independence (most often used versions assume fully independence)

Lemma 2 (Strengthening of Marcinkiewicz-Zygmund's Inequality for $p = 4$). *Suppose that $\{\xi(x)\}_{x\in\mathcal{X}}$ are 4-wise independent, with zero mean. Then* *we have*

$$\frac{1}{\sqrt{3}}\left(\sum_{x\in\mathcal{X}}\mathrm{Var}(\xi(x))\right)^{\frac{1}{2}} \leqslant \mathbb{E}\left|\sum_{x\in\mathcal{X}}\xi(x)\right| \leqslant \left(\sum_{x\in\mathcal{X}}\mathrm{Var}(\xi(x))\right)^{\frac{1}{2}}$$

$$\mathbb{E}\left|\sum_{x\in\mathcal{X}}\xi(x)\right|^{2} = \sum_{x\in\mathcal{X}}\mathrm{Var}(\xi(x))$$

$$\left(\sum_{x\in\mathcal{X}}\mathrm{Var}(\xi(x))\right)^{2} \leqslant \mathbb{E}\left|\sum_{x\in\mathcal{X}}\xi(x)\right|^{4} \leqslant 3\left(\sum_{x\in\mathcal{X}}\mathrm{Var}(\xi(x))\right)^{2}$$

The proof appears in Appendix A.1.

2.3 Anticontentration bounds

Lemma 3 (Paley-Zygmund Inequality). *For any positive random variable Z and a parameter $\theta\in(0,1)$ we have*

$$\Pr\left[Z > \theta\,\mathbb{E}\,Z\right] \geqslant (1-\theta)^2\frac{(\mathbb{E}\,Z)^2}{\mathbb{E}\,Z^2}.$$

By applying Lemma 3 to the setting of Lemma 2, and choosing $\theta = \frac{1}{\sqrt{3}}$ we obtain

Corollary 3 (Anticoncentration for walks with 4-wise independent increments). *Suppose that $\{\xi(x)\}_{x\in\mathcal{X}}$ are 4-wise independent, then we have*

$$\Pr\left[\left|\sum\xi(x)\right| > \frac{1}{3}\left(\sum\mathrm{Var}(\xi(x))\right)^{\frac{1}{2}}\right] > \frac{1}{17}.$$

where the summation is over $x\in\mathcal{X}$.

3 Proof of Theorem 1

Lemma 4 (Characterizing smooth min-entropy). *For any random variable X with values in a finite set $\mathcal{X}$, any δ and k we have the following equivalence*

$$H^{\delta}_{\infty}(X) \geqslant k \iff \sum_{x\in\mathcal{X}}\max\left(P_X(x)-2^{-k},0\right) \leqslant \delta.$$

The proof appears in Appendix B.1. We will work with the following equivalent statement

Corollary 4 (No smooth min-entropy k implies bias w.r.t. distributions of min-entropy k, over at most 2^k elements). *We have $H^{\delta}_{\infty}(X) < k$ if and only if there exists a set S of at most 2^k elements such that*

$$\sum_{x\in S}|P_X(x)-P_Y(x)| > \delta$$

for all Y of min-entropy at least k.

Proof (Proof of Corollary 4). The direction $\Longleftarrow$ trivially follows by the definition of smooth min-entropy. Now assume $H_{\infty}^{\delta}(X) < k$, then by Lemma 4 we have $\sum_{x\in S}\max\left(P_X(x) - 2^{-k}\right) > \delta$ where S is the set of all x such tat $P_x(x) > 2^{-k}$. In particular

$$\sum_{x\in S}\max\left(P_X(x) - P_Y(x)\right) > \delta$$

for any Y of min-entropy k, because then $P_Y(x) \leqslant 2^{-k}$. Moreover, by the definition of S it follows that $|S| < 2^k$.

Lemma 5 (Bias implies euclidean distance). *For any (not necessarily probabilistic) measures P_X, P_Y on $\mathcal{X}$ and any subset S of $\mathcal{X}$ we have*

$$\left(\sum_{x\in S}\left(P_X(x) - P_Y(x)\right)^2\right)^{\frac{1}{2}} > |S|^{-1}\sum_{x\in S}\left|P_X(x) - P_Y(x)\right|.$$

Proof. By the Jensen Inequality we have $|S|^{-1}\left(\sum_{x\in S}\left(P_X(x) - P_Y(x)\right)^2\right) > \left(|S|^{-1}\sum_{x\in S}|P_X(x) - P_Y(x)|\right)^2$ which is equivalent to the statement.

Corollary 5 (No smooth min-entropy implies euclidean distance to min-entropy distributions). *Suppose that $H_{\infty}^{\delta}(X) < k$. Then for any Y of min-entropy at least k we have $\left(\sum_x |P_X(x) - P_Y(x)|^2\right)^{\frac{1}{2}} > 2^{-\frac{k}{2}}\delta$.*

Proof (Proof of Corollary 5). It suffices to combine Lemma 5 and Corollary 4.

By Corollary 3 we conclude that the advantage of a random distinguisher for any two measures (in our case P_X and P_Y) equals the euclidean distance.

Lemma 6 (The advantage of a random distinguisher equals the euclidean distance). *Let $\{\mathsf{D}(x)\}_{x\in\{0,1\}^n}$ be 4-wise independent as indexed by x and such that $\mathsf{D}(x)$ outputs a random element from $\{-1,1\}$. Then for any measures μ_1 and μ_2 and any set S we have*

$$\left|\sum_{x\in S}\mathsf{D}(x)(P_X(x) - P_Y(x))\right| > \frac{1}{3}\cdot d_2(P_X; P_Y)$$

with probability $\frac{1}{17}$ over the choice of D (the result holds for any measures in place of P_X, P_Y).

For our case, that is the setting in Lemma 5, we obtain

Corollary 6 (A random attack achieves $\Omega\left(2^{-k}\delta\right)$). *For X, Y as in Corollary 5, and D as in Lemma 6 we have $\mathsf{Adv}^{\mathsf{D}}(X;Y) = \frac{1}{3}\cdot 2^{-\frac{k}{2}}$ with probability $\frac{1}{17}$ over D.*

Partitioning the domain into T slices Let $h : \{0,1\}^n \to [1 \ldots 2^t]$, where $t = \lceil \log T \rceil$, be a 8-universal hash function (we will see that the order 8 is important here!). Define $S_i = \{x : h(x) = i\}$, $\Delta(x) = P_X(x) - P_Y(x)$ and consider advantages on slices S_i

$$\mathsf{Adv}^{\mathsf{D}}(S_i) = \sum_{x \in S_i} \Delta(x) \mathsf{D}(x) \mathbf{1}_{S_i}(x)$$

The following corollary shows that on each of our T slices, we get the advantage $|T|^{-\frac{1}{2}} 2^{-\frac{k}{2}} \delta$. The proof appears in Appendix C.1.

Corollary 7 ((Mixed) moments of slice advantages). *For* D*,* $\{S_u\}_u$ *as above and every* i, j

$$\mathbb{E}_{\mathsf{D},\{S_u\}_u} \mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) \geqslant 3^{-\frac{1}{2}} T^{-\frac{1}{2}} \cdot d_2(P_X; P_Y)$$
$$\mathbb{E}_{\mathsf{D},\{S_u\}_u} \left(\mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) \mathsf{Adv}^{\mathsf{D}}_{S_j}(X;Y) \right) \leqslant T^{-1} \cdot d_2(P_X; P_Y)^2$$

(the statement is valid for arbitrary measures in place of P_X, P_Y*).*

Denote $Z = \sum_i \mathsf{Adv}^{\mathsf{D}}(S_i)$. Using Lemma 3 with $\theta = \frac{1}{\sqrt{3}}$ where we compute $\mathbb{E}\, Z^2$ and $\mathbb{E}\, Z$ according to Corollary 7 we obtain $\Pr\left[|Z| > \frac{1}{\sqrt{3}} \cdot \mathbb{E}\, |Z| \right] \geqslant \frac{1}{17}$. Bounding once again $\mathbb{E}\, |Z|$ as in Corollary 7 we get

Corollary 8 (Total advantage on all parition slices). *For* X, Y *as in Corollary 5,* D *and* S_i *defined above we have*

$$\Pr_{\mathsf{D},\{S_u\}_u} \left[\sum_{i=1}^{T} \mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) \geqslant \frac{1}{3} \cdot T^{\frac{1}{2}} 2^{-\frac{k}{2}} \delta \right] \geqslant \frac{1}{17}.$$

(for general X, Y *the lower bound is* $\Omega(1) \cdot T^{\frac{1}{2}} \cdot d_2(P_X; P_Y)$*).*

The corollary shows that the *total absolute advantage* over all partition slices, is as expected. Since $\{S_i\}_i$ is a partition we have

$$\sum_{i=1}^{T} \mathsf{Adv}^{\mathsf{D}}_{S_i}(X;Y) = \sum_{i=1}^{T} \left| \sum_{x \in S_i} (P_X(x) - P_Y(x)) \mathsf{D}(x) \right|$$
$$= \sum_{x} (P_X(x) - P_Y(x)) \mathsf{D}(x) \beta_i$$

where $\beta_i = \operatorname{sgn}\left(\sum_{x \in S_i} (P_X(x) - P_Y(x)) \mathsf{D}(x) \right)$ is a sign of the advantage of D restricted to the i-th slice. This shows that by "flipping" the distinguisher output on slices we achieve the sum of individual advantages. Thinking of bits β_i as extra advice we obtain

Corollary 9 (Computing total advantage by one distinguisher). *For* X, Y *as in Corollary 5,* D *and* $\{S_i\}_i$ *defined above there exists a modification to* D *which in time* $O(n)$ *and advice* $O(T)$ *achieves advantage* $\frac{1}{3} \cdot T^{\frac{1}{2}} 2^{-\frac{k}{2}} \delta$ *with probability* $\frac{1}{17}$*.*

Finally by setting $\epsilon = T^{\frac{1}{2}} 2^{-\frac{k}{2}} \delta$ and manipulating T we arrive at

Corollary 10 (Continue tradeoff). *For any ϵ there exists T such that the distinguisher in Corollary 9 has advantage ϵ and circuit complexity $s = O\left(2^k \epsilon^2 \delta^{-2}\right)$.*

References

Ber97. Bonnie Berger, *The fourth moment method*, SIAM J. Comput. **26** (1997), no. 4, 1188–1207.

DTT10. Anindya De, Luca Trevisan, and Madhur Tulsiani, *Time space tradeoffs for attacks against one-way functions and prgs*, Advances in Cryptology - CRYPTO 2010, 30th Annual Cryptology Conference, Santa Barbara, CA, USA, August 15-19, 2010. Proceedings, 2010, pp. 649–665.

Khi24. Aleksandr Khintchine, *über einen satz der wahrscheinlichkeitsrechnung*, Fundamenta Mathematicae **6** (1924), no. 1, 9–20.

MZ37. J Marcinkiewicz and A Zygmund, *Quelques théoremes sur les fonctions indépendantes'*, Fund. Math **29** (1937), 60–90.

SSS93. Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan, *Chernoff-hoeffding bounds for applications with limited independence*, Proceedings of the Fourth Annual ACM/SIGACT-SIAM Symposium on Discrete Algorithms, 25-27 January 1993, Austin, Texas., 1993, pp. 331–340.

A Random Walks

A.1 Proof of Lemma 2

Denote $Z = \sum_x \xi(x)$. Since $\xi(x)$ are (in particular) 2-wise independent with zero mean, we get

$$\mathbb{E}\left(\sum_x \xi(x)\right)^2 = \sum_{x,y} \mathbb{E}\left(\xi(x)\xi(y)\right) = \sum_{x=y} \mathbb{E}\left(\xi(x)\xi(y)\right) = \sum_x \mathrm{Var}(\xi(x)).$$

The fourth moment is somewhat more complicated

$$\begin{aligned}
\mathbb{E}\left(\sum_x \xi(x)\right)^4 &= \sum_{x_1,x_2,x_3,x_4} \mathbb{E}\left(\xi(x_1)\xi(x_2)\xi(x_3)\xi(x_4)\right) \\
&= \sum_{x_1=x_2=x_3=x_4} \mathbb{E}\left(\xi(x_1)\xi(x_2)\xi(x_3)\xi(x_4)\right) + \\
&\quad + 3 \sum_{x_1=x_2\neq x_3=x_4} \mathbb{E}\left(\xi(x_1)\xi(x_2)\xi(x_3)\xi(x_4)\right) \\
&= \sum_x \mathbb{E}\,\xi(x)^4 + 3\sum_{x\neq y} \mathbb{E}\,\xi(x)^2\,\mathbb{E}\,\xi(y)^2 \\
&= 3\left(\sum_x \xi(x)^2\right)^2 - 2\sum_x \xi(x)^4
\end{aligned}$$

(whenever $\xi(x)$ appears in an odd power, for example $x = x_1 \neq x_2 = x_3 = x_4$ the expectation is zero. This observation may be used to simplify and bound also higher moments, see [SSS93]). It remains to estimate the first moment. By Corollary 2 and bounds on the second and fourth moment we have just computed we obtain

$$\frac{1}{\sqrt{3}} \cdot \left(\sum_{x \in \mathcal{X}} \operatorname{Var}(\xi(x)) \right)^{\frac{1}{2}} \leqslant \mathbb{E} \left| \sum_{x \in \mathcal{X}} \xi(x) \right|$$

and the upper bound follows by Jensen's Inequality (with constant 1).

B Smooth Min-Entropy

B.1 Proof of Lemma 4

Suppose that $H_\infty^\delta(X) \geqslant k$. then, by definition, there is Y such that $H_\infty(Y) \geqslant k$ and $\sum_{x: P_X(x) > P_Y(x)} P_X(x) - P_Y(x) \leqslant \delta$. Since all the summands are positive and since $P_Y(x) \leqslant 2^{-k}$, ignoring those x for which $P_Y(x) < 2^{-k}$ yields

$$\sum_{x: P_X(x) > 2^{-k}} P_X(x) - P_Y(x) \leqslant \delta.$$

Again, since $P_Y(x) \leqslant 2^{-k}$ we obtain

$$\sum_{x: P_X(x) > 2^{-k}} P_X(x) - 2^{-k} \leqslant \delta,$$

which finishes the proof of the "$\Longrightarrow$" part.

Assume now that $\delta' = \sum_{x \in \mathcal{X}} \max\left(P_X(x) - 2^{-k}, 0\right) \leqslant \delta$. Note that

$$\sum_{x \in \mathcal{X}} \max\left(P_X(x) - \frac{1}{2^k}, 0\right) + \sum_{x \in \mathcal{X}} \max\left(\frac{1}{2^k} - P_X(x), 0\right) =$$
$$= 2 \sum_{x \in \mathcal{X}} \left| P_X(x) - \frac{1}{2^k} \right| \geqslant 2 \sum_{x \in \mathcal{X}} \max\left(P_X(x) - \frac{1}{2^k}, 0\right)$$

and therefore we have $\sum_{x \in \mathcal{X}} \max\left(2^{-k} - P_X(x), 0\right) \geqslant \delta'$. By this observation we can construct a distribution Y by shifting δ' of the mass of P_X from the set $S^- = \{x : P_X(x) > 2^{-k}\}$ to the set $\{x : 2^{-k} \geqslant P_X(x)\}$ in such a way that we have $P_Y(x) \leqslant 2^{-k}$ for all x. Thus $H_\infty(Y) \geqslant k$ and since a δ' fraction of the mass is shifted and redistributed we have $d_1(X; Y) \leqslant \delta'$. This finishes the proof of the "$\Longleftarrow$" part.

C Claims in the proof of Theorem 1

C.1 Proof of Corollary 7

For simplicity denote $\varDelta(x) = P_X(x) - P_Y(x)$. Define

$$\begin{aligned} E &= \mathbb{E}\left(\mathsf{Adv}^{\mathsf{D}}\left(S_i\right) \cdot \mathsf{Adv}^{\mathsf{D}}\left(S_j\right)\right) \\ &= \mathbb{E}\left|\sum_{x,x'} \varDelta(x)\varDelta(x')\mathbf{1}_{S_i}(x)\mathbf{1}_{S_j}(x')\mathsf{D}(x)\mathsf{D}(x')\right| \end{aligned}$$

Note that $\mathbf{1}_{S_i}(x)\mathbf{1}_{S_j}(x')\mathsf{D}(x)\mathsf{D}(x')$, understood as a family indexed by x and x', are 2-wise independent and have zero-mean. Therefore, by Lemma 2 for $i \neq j$

$$\begin{aligned} E &\leqslant \left(\sum_{x,x'} \mathrm{Var}(\varDelta(x)\varDelta(x')\mathbf{1}_{S_i}(x)\mathbf{1}_{S_j}(x')\mathsf{D}(x)\mathsf{D}(x'))\right)^{\frac{1}{2}} \\ &= \left(T^{-2}\sum_{x\neq x'} \varDelta(x)^2\varDelta(x')^2\right)^{\frac{1}{2}}, \end{aligned}$$

which means that

$$\mathbb{E}\left(\mathsf{Adv}^{\mathsf{D}}\left(S_i\right) \cdot \mathsf{Adv}^{\mathsf{D}}\left(S_j\right)\right) \leqslant T^{-1} \cdot \left(\left(\sum_x \varDelta(x)^2\right)^2 - \sum_x \varDelta(x)^4\right)^{\frac{1}{2}}.$$

In turn, again by Lemma 2, we have

$$\mathbb{E}\left(\mathsf{Adv}^{\mathsf{D}}\left(S_i\right)\right)^2 = T^{-1} \cdot \sum_x \varDelta(x)^2.$$

It follows that for any i, j we get

$$\mathbb{E}\left(\mathsf{Adv}^{\mathsf{D}}\left(S_i\right)\mathsf{Adv}^{\mathsf{D}}\left(S_j\right)\right) \leqslant T^{-1} \cdot \sum_x \varDelta(x)^2,$$

which proves the second inequality in the corollary. Finally, note that by Lemma 2 we also have

$$\mathbb{E}\,\mathsf{Adv}^{\mathsf{D}}\left(S_i\right) \geqslant 3^{-\frac{1}{2}}\left(\sum_x \varDelta(x)^2\right)^{\frac{1}{2}}$$

which is the first inequality claimed in the corollary.