A Comprehensive Comparison of Shannon Entropy and Smooth Renyi Entropy

November 29, 2014

Abstract

We provide a new result that links two crucial entropy notions: Shannon Entropy H_1 and collision entropy H_2. Our formula gives the *worst possible* amount of collision entropy in a probability distribution, when its Shannon Entropy is fixed.

Our results and techniques used in the proof immediately imply many quantitatively tight separations between Shannon and smooth Renyi entropy, which were previously known as qualitative statements or one-sided bounds. In particular, we precisely calculate the number of bits that can be extracted from a Shannon Entropy source, and calculate how far from the uniform distribution is a distribution with the given amount of Shannon Entropy. To illustrate our results we provide clear numerical examples.

In the typical situation, when the gap between Shannon Entropy of a distribution and its length is bigger than 1, the length of the extracted sequence is very small, even if we allow the randomness quality to be poor. In the case of almost full entropy, where the gap is close to 0, the ℓ_2-distance to uniform is roughly of the same order as the gap. Therefore, it is actually not possible to decide the strong quality of supposed true randomness, efficiently and at extremely high confidence level, by means of Shannon Entropy estimators, like Maurer's Universal Test or others.

Our approach involves convex optimization techniques, applied to characterize worst case distributions, and the use of the Lambert W function, by which we resolve equations coming from Shannon Entropy constraints. We believe that it may be useful and of independent interests elsewhere, particularly for studying Shannon Entropy with constraints.

Keywords. Renyi Entropy, Smooth Entropy, Entropy Estimators, Convex Optimization, Lambert W Function

1 Introduction

1.1 Entropy Measures

Entropy, as a measure of randomness contained in a probability distribution, is the fundamental concept in information theory and cryptography. There exists many entropy definitions and they are not equally good for all applications. While the most famous (and most liberal) Shannon Entropy [Sha48] is extremely useful in information theory, the use of much more conservative measures, like min-entropy or collision entropy, is necessary in cryptographic applications, like extracting randomness. Misunderstanding about what is a suitable entropy notion, is a common problem in practical designs and not only of a theoretical concern because it leads to vulnerabilities due to overestimating security. In fact, when entropy is underestimated, security of real-world

applications can be broken [DPR+13]. That's why the standards for random bits generating [BK12] strongly recommend the use of min-entropy for secure implementations.

However, under some circumstances it is possible to relate the Shannon Entropy and the amount of extractable entropy. Basically, this is when the entropy source generates bits or blocks of bits in an independent way. Such an assumption (even if idealistic) is of a crucial importance for provable secure analysis of true random number generators [BL05, BKMS09, VSH11, LPR11], one of the fundamental and most challenging topics in real-world cryptography.

1.2 Our Results and Techniques

1.2.1 Brief Summary.

We investigate in deep details the gap between Shannon Entropy and Renyi Entropy (focusing on smooth collision entropy and smooth min-entropy) in a given entropy source. We impose no restrictions on the source and obtain general and tight bounds as well as identify worst cases. Our results are mostly negative, in the sense that the gap might be extremely big and even almost full Shannon Entropy does not guarantee the closeness to the uniform distribution. Our negative results are partially known in the literature or in folklore. However, to the best of our knowledge, our analysis for the first time provides a comprehensive and detailed study of this problem, establishing really tight bounds. Moreover, it may be of independent interest because of the techniques we successfully applied.

1.2.2 Results.

Bounding Renyi Entropy by Shannon Entropy. Interested in establishing a bound on the amount of extractable entropy in terms of Shannon Entropy only, we ask the following question

> **Q**: Suppose that the Shannon Entropy $\mathrm{H}_1(X)$ of a random variable $X \in \{0,1\}^n$ is at least k. What is the best possible lower bound on the collision entropy $\mathrm{H}_2(X)$?

Our Theorem 1 gives a complete answer to this question. The conclusion has been briefly summarized in the table below

Domain of X	$\mathrm{H}_1(X)$	Region	Max. ℓ_2-distance to uniform	Min. value of $\mathrm{H}_2(X)$
$\{0,1\}^n$	$n-\Delta$	$2^n\Delta \geqslant 13$	$\Theta\left(\frac{\Delta}{\log(2^n\Delta)}\right)$	$n-\log_2\left(1+\Theta\left(2^n\Delta^2\log^{-2}(2^n\Delta)\right)\right)$
		$2^n\Delta \leqslant 13$	$O\left(\Delta\right)$	$n-\log_2\left(1+O\left(2^n\Delta^2\right)\right)$

Table 1: Minimal collision entropy given Shannon Entropy constraints

The statement and some further discussions can be found in Section 3. Interestingly, the shape of the "worst" distribution X is pretty simple: a combination of a one-point heavy mass with a flat distribution outside. In fact, it has been already observed in the literature that such a shape provides good separations for Shannon Entropy [Cac97]. But to our knowledge, we first prove the opposite: this shape is really best possible.

Infeasibility of uniformity tests based on entropy estimators. If an n-bit random variable X satisfies $\mathrm{H}_1(X) = n$ then it must be uniform. It might be tempting then to think that the very small entropy gap $\Delta = n - \mathrm{H}_1(X)$ (when entropy is extremely "condensed") implies the closeness to the uniform distribution.

> **Q**: Suppose that the Shannon Entropy $\mathrm{H}_1(X)$ of a random variable $X \in \{0,1\}^n$ is at least $n - \Delta$ where $\Delta \approx 0$. What is the best possible upper bound on the distance between X and the uniform distribution U_n?

Using Theorem 1 we prove that for the statistical distance (ℓ_1 distance) the gap Δ can be as small as ϵ but still the source is ϵ/n-far from the uniform distribution. This shows that an application of entropy estimators to test sequences of truly random bits might be problematic, because estimating entropy within an additive error smaller than negligible value ϵ is computationally inefficient. Having said this, we stress that entropy estimators like Maurer's Universal Test [Mau92] are still powerful tools of discovering the most of defects, which appear within a broader margin of error. See Corollary 3 and Remark 4 for the statement and a short discussion.

Large gap between Shannon and smooth collision entropy. The collision entropy of a distribution X constitutes a lower bound on the number of extractatable almost-uniform bits. Therefore, the following question is natural

> **Q**: Suppose that the Shannon Entropy $\mathrm{H}_1(X)$ of a random variable $X \in \{0,1\}^n$ is at least $n - \Delta$ where $\Delta \leqslant 1$. What is the best possible lower bound on $\mathrm{H}_2(X)$? Does it help if we relax the problem and consider $\mathrm{H}_2(X')$ where X' is close to X?

As a negative result, we demonstrate that the gap between the Shannon Entropy and Renyi Entropy could be almost as big as the length of the entropy source output (that is almost maximal possible). Moreover, smoothing entropy, even with weakly security, does not help. For example, we construct a 256-bit string of more than 255 bits of Shannon Entropy, but only 19 bits of (smooth) Renyi entropy. For more details and the precise statement we refer to Corollary 4 in Section 4.2. To our knowledge, our analysis of smooth Renyi entropy is original, though the separation for non-smooth entropy is known [BBM95]. The separation is an easy corollary from the proof of Theorem 1.

Large gap between Shannon and extractable Entropy. Min entropy gives only a lower bound on extractable entropy. However, its smooth version can be used to establish an upper bound on the amount of almost random bits, of required quality, that can be extracted from a given source.

> **Q**: Suppose that the Shannon Entropy $\mathrm{H}_1(X)$ of a random variable $X \in \{0,1\}^n$ is at least $n - \Delta$ where $\Delta < 1$. How many bits that are close to uniform can be extracted from X?

Again, analogously to the previous result, we provide a separation between Shannon and extractable entropy, where the gap is almost as big as the length of the random variable. For example, we construct a 256-bit string of more than 255.5 bits of Shannon Entropy, but only 10 bits of extractable entropy, even if we allow them to be of very weak quality, not really close to uniform! For more details and the precise statement we refer to Corollary 5 in Section 4.2. To our knowledge, the concrete tight bounds we provide are new, though a similar "extreme" numerical example can be found in [Cac97]. The separation is again a straightforward application of ideas behind Theorem 1.

Converting Shannon Entropy into Renyi Entropy. Even though the gap in our separations are almost as big as the length of the source output, there might be some small amount of Renyi Entropy present in every distribution of high Shannon Entropy.

> **Q**: Suppose that the Shannon Entropy of a random variable $X \in \{0,1\}^n$ is at least $n - \Delta$ where $\Delta \geqslant 1$. Does X have some non-trivial amount of collision entropy?

Using our Theorem 1 we establish a simple and tight bound on this amount: it is about $2\log_2 n - 2\log_2 \Delta$. For example, in the concrete case of a 256-bit string of Shannon Entropy 255 we find that the necessary amount of Renyi entropy is 15. We also establish an interesting rule of thumb: for much more than one bit of Renyi entropy in the output of a source, its Shannon Entropy must be bigger than the half of its length. The more details can be found in Corollary 6 in Section 4.2. Our conversion can be applied in some settings where Shannon Entropy is the easiest or most reliable entropy notion to measure, but for security reasons a more conservative measure of randomness is preferable.

1.2.3 Techniques

To prove our main technical results, we use standard convex optimization techniques combined with some calculus which allows us to deal with implicit equations. In particular, we demonstrate that the Lambert-W function is useful in studying Shannon Entropy constraints.

1.3 Organization of the paper

We start with necessary definitions and explanations of basic concepts in Section 2. Our main result is discussed in Section 3. Further applications are given in Section 4. We end with the conclusion in Section 5. The proofs of main results, which are technical and complicated a bit, appear in Section 5.

2 Preliminaries

2.1 Basic Notions.

By U_S we denote the uniform distribution over a set S, and U_n is a shortcut for the uniform n-bit distribution. The closeness of two distributions X, Y over the same domain is most commonly measured by the so called statistical or variational distance $\mathrm{SD}(X;Y)$. It is defined as the half of the ℓ_1-distance between the probability mass functions $\mathrm{SD}(X;Y) = \frac{1}{2}\sum_x |\Pr[X=x] - \Pr[Y=x]|$. In this paper we use also the ℓ_2-distance between probability distributions, defined as $\mathrm{d}_2(X;Y) = \sqrt{\sum_x (\Pr[X=x] - \Pr[Y=x])^2}$. For convenience we define also the collision probability of X as the probability that two independent copies of X collide: $\mathrm{CP}(X) = \sum_x \Pr[X=x]^2$.

2.2 Entropy Definitions.

Below we define the three key entropy measures, already mentioned in the introduction. It is worth of noting that they all are special cases of a much bigger parametrized family of Renyi entropies. However the common convention in cryptography, where only these three matter, is to slightly abuse the terminology and to refer to collision entropy when talking about Renyi entropy, keeping the names for Shannon and Min-Entropy.

Definition 1 (Entropy notions)**.** *The Shannon Entropy* $\mathrm{H}(X) = \mathrm{H}_1(X)$*, the collision entropy (or*

Renyi entropy) $\mathrm{H}_2(X)$*, and the Min-Entropy* $\mathrm{H}_\infty(X)$ *of a distribution* X *are defined as follows*

$$\mathrm{H}(X) = \sum_x \Pr[X = x] \log \Pr[X = x] \tag{1}$$

$$\mathrm{H}_2(X) = -\log \left(\sum_x \Pr[X = x]^2 \right) \tag{2}$$

$$\mathrm{H}_\infty(X) = -\log \max_x \Pr[X = x]. \tag{3}$$

Remark 1 (Comparing different entropies)**.** *It is easy to see that we have*

$$\mathrm{H}(X) \geqslant \mathrm{H}_2(X) \geqslant \mathrm{H}_\infty(X),$$

with the equality if and only if X *is uniform.*

2.3 Entropy Smoothing

The Concept. Entropy Smoothing is a very useful concept of replacing one distribution by a distribution which is very close in the statistical distance (which allows keeping its most important properties, like the amount of extractable entropy) but more convenient for the application at hand (e.g. a better structure, removed singularities, more entropy).

Applications. The smoothing technique is typically used to *increase entropy* by cutting of big but rare "peaks" in a probability distribution. Probably the most famous example is the so called Asymptotic Equipartition Property (AEP). Imagine a sequence X of n independent Bernoulli trials, where 1 appears with probability $p > 1/2$. Among the all n-bit sequences the most likely ones are those with 1 on almost all places. In particular $\mathrm{H}_\infty(X) = -n \log p$. However, for the most of sequences the number of 1's oscillates around pn (these are so called typical sequences). By Hoeffding's concentration inequality, the number of 1's is at most $pn + h$ with probability $1 - \exp(-2h^2/n)$. For large n and suitably chosen h, the distribution of X approaches a distribution X' of min-entropy $\mathrm{H}_\infty(X') \approx -n(p \log p + (1-p)\log(1-p)) \approx \mathrm{H}(X)$ (the relative error here is of order $O(n^{-1/2})$), much larger than the min-entropy of the original distribution! A quantitative version of this fact was used in the famous construction of a pseudorandom generator from any one-way function [HILL88]. Renner and Wolf revisited the smoothing technique in entropy framework and came up with new applications [RW04].

Definition 2 (Smooth Entropy, [RW04])**.** *Suppose that* $\alpha \in \{1, 2, \infty\}$*. We say that the* ϵ*-smooth entropy of order* α *of* X *is at least* k *if there exists a random variable* X' *such that* $\mathrm{SD}(X; X') \leqslant \epsilon$ *and* $\mathrm{H}_\alpha(X') \geqslant k$*.*

For shortness, we also say smooth Shannon Entropy, smooth Renyi entropy or smooth min-entropy. We also define the *extractable entropy* of X as follows

Definition 3 (Extractable Entropy, [RW05])**.** *The* ϵ*-extractable entropy of* X *is defined to be*

$$\mathrm{H}^\epsilon_{\mathrm{ext}}(X) = \max_{\mathcal{U}:\ \exists f \in \Gamma^\epsilon(X \to \mathcal{U})} \log |\mathcal{U}| \tag{4}$$

where $\Gamma^\epsilon(X \to \mathcal{U})$ *consists of all functions* f *such that* $\mathrm{SD}(f(X, R); U_{\mathcal{U}}, R) \leqslant \epsilon$ *where* R *is uniform and independent of* X *and* $U_{\mathcal{U}}$ *.*

2.4 Extractors.

Roughly speaking, an extractor is a randomized function which produces an almost uniform string from a longer string but not of full entropy. The randomization here is necessary if one wants an extractor working for high-entropy sources; the role of that auxiliary randomness is similar to the purpose of catalysts in chemistry.

Definition 4 (Strong Extractors [NZ96]). *A strong (k, ϵ)-extractor is a function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^k$ *such that*

$$\mathrm{SD}(\mathrm{Ext}(X, U_d); U_{k+d}) \leqslant \epsilon. \tag{5}$$

A very simple, efficient and optimal (regarding to the necessarily entropy loss) extractor is based on universal hash functions. Recall that a class $\mathcal{H}$ of functions from n to m bits is universal [CW79] if for any different x, y there are exactly $|\mathcal{H}|/2^m$ functions $h \in \mathcal{H}$ such that $h(x) = h(y)$.

Lemma 1 (Leftover Hash Lemma). *Let $\mathcal{H}$ be a universal class of functions from n to m bits, let H be chosen from $\mathcal{H}$ at random and let X be an n-bit random variable. If $\mathrm{H}_2(X) \geqslant k$, then $\mathrm{SD}(H(X), H; U_m, H) \leqslant \frac{1}{2} \cdot 2^{\frac{m-k}{2}}$.*

By Lemma 1 and the properties of the statistical distance we obtain

Corollary 1 (Bound on extractable entropy, [RW05]). *We have $\mathrm{H}^{\epsilon}_{\mathrm{ext}}(X) \geqslant \mathrm{H}_2^{\epsilon/2}(X) - 2\log(1/\epsilon) - 1$.*

Note that to extract k bits ϵ-close to uniform we need to invest $k + 2\log(1/\epsilon)$ bits of entropy; the loss of $2\log(1/\epsilon)$ bits here is optimal [RTS00]. Wile there are many other extractors, the Leftover Hash Lemma is particularly often used in the TRNG design [BST03, BKMS09, VSH11] because it is simple, efficient, and provable secure. Extractors based on the LHL are also very important in key derivation problems [BDK+11].

3 Main Result

3.1 Maximizing Collisions given Shannon Entropy

Below we answer the posted question on the best bound on H_2 in terms of H_1. The "worst case" distribution, which minimizes the gap, is pretty simple: it is a combination of a one-point mass at some point and a uniform distribution outside.

Theorem 1. *Let X be a random variable with values in a d-element set. If $\mathrm{H}(X) = k$, then*

$$\mathrm{H}_2(X) \geqslant -\log_2 \left(b^2 + \frac{(1-b)^2}{d-1} \right) \tag{6}$$

where b is the unique solution to

$$\mathrm{H}(b) + (1-b)\log_2(d-1) = k \tag{7}$$

under the restriction $b \geqslant \frac{1}{d}$ ($H(b)$ denotes the entropy of a bit equal 1 with probability b). The bound in Equation (6) is best possible.

Remark 2 (The implicit equation in Theorem 1). *The number b is defined nondirectly depending on d and k. In Section 3.2, we will show how to accurately approximate the solution of this equation.*

The proof of Theorem 1 appears in Appendix A. The main idea is to write down the posted question as a constrained optimization problem and apply standard Lagrange multipliers techniques.

3.2 Closed-Form Bounds for Solutions

Below we present a tight formula approximating the solution to Equation (7). We will substitute it to Equation (6) in order to obtain a closed-form expression.

Lemma 2 (The solution for moderate gaps)**.** *Let b be the solution to Equation (7) and let $\Delta = \log_2 d - k$ be the entropy gap. Suppose $d\Delta \geqslant 13$. Then we have*

$$\frac{0.84\Delta}{\log_2(d\Delta) - 1.52} \leqslant b \leqslant \frac{1.37\Delta}{\log_2(d\Delta) - 1.98} \tag{8}$$

The proof is referred to Appendix B. The main idea is to solve Equation (8) approximately using the so called Lambert W function, that matches Shannon-like expressions of the form $y \log y$. Here we discuss the lemma and its applications.

Remark 3 (Establishing tighter constants)**.** *The numerical constants in Lemma 2 can be made sharper if needed. Under the (mild) assumption that $\Delta^{-1} = 2^{o(\log_2 d)}$ and $d\Delta = \omega(1)$ (as d grows), one can get*

$$b = \frac{(1 + o(1))\Delta}{\log_2(d\Delta) - \log_2 \mathrm{e} - \log_2 \log_2 \mathrm{e} + o(1)} \tag{9}$$

The gap between 1.52 and 1.98 is self-improving, in the sense that knowing in advance a better upper bound on b one makes it closer to 0. In turn, the gap between 0.84 and 1.37 can be made closer to 0 by choosing in the proof a more accurate approximation for the Lambert W function.

Now we are ready to compute minimal collision entropy given Shannon Entropy.

Corollary 2 (Minimal collision entropy, general case)**.** *Let X^* minimizes $\mathrm{H}_2(X)$ subject to $\mathrm{H}(X) \geqslant n - \Delta$ where X takes its values in a given d-element set. If $d\Delta \geqslant 13$ then*

$$\frac{0.55\Delta}{\log_2(d\Delta)} \leqslant \mathrm{d}_2(X^*; U) \leqslant \frac{3.24\Delta}{\log_2(d\Delta)}, \tag{10}$$

where U is uniform over the domain of X. If $d\Delta < 13$ then

$$\mathrm{d}_2(X^*; U) < 0.88\Delta. \tag{11}$$

The collision entropy is obtained as $\mathrm{H}_2(X^*) = -\log_2\left(\frac{1}{d} + \mathrm{d}_2(X^*; U)^2\right)$.

Proof of Corollary 2. We will consider two cases.
Case I: $d\Delta \geqslant 13$. By Lemma 2 we get

$$\frac{0.84\Delta}{\log_2(d\Delta)} \leqslant b \leqslant \frac{2.95\Delta}{\log_2(d\Delta)} \tag{12}$$

By the last inequality and the fact that $x \to \frac{x}{\log_2 x}$ is increasing for $x \geqslant \mathrm{e}$ we get

$$bd \geqslant \frac{0.84 d\Delta}{\log_2(d\Delta)} \geqslant 2.95$$

Let $b_0 = \frac{1}{d}$. By the last inequality we get $b - b_0 \geqslant 0.66b$. Since

$$b^2 + \frac{(1-b)^2}{d-1} = b_0 + \frac{d}{d-1} \cdot (b - b_0)^2,$$

by the identity $\mathrm{d}_2(X;U)^2 = \sum_x \Pr[X=x]^2 - \frac{1}{d}$ and the definition of collision entropy we get

$$\mathrm{d}_2(X^*,U)^2 = \mathrm{CP}(X^*) - b_0 = \frac{d}{d-1}\cdot(b-b_0)^2.$$

Note that $d\Delta \geqslant 13$ implies $d\log_2 d \geqslant 13$ (because $\Delta \leqslant \log_2 d$) and hence $d > 5$. By this inequality and $b - b_0 \geqslant 0.66b$ we finally obtain

$$0.43b^2 \leqslant \mathrm{d}_2(X^*;U)^2 \leqslant 1.2b^2 \tag{13}$$

and the result for the case $d\Delta \geqslant 13$ follows by combining Equation (12) and Equation (13).

Case II: $d\Delta < 13$. We do a trick to "embed" our problem into a higher dimension. If $\boldsymbol{p} \in \mathbb{R}^d$ is the distribution of X, define $\boldsymbol{p}' \in \mathbb{R}^{d+1}$ by $\boldsymbol{p}'_i = (1-\gamma)\boldsymbol{p}'_i$ for $i \leqslant d$ and $\boldsymbol{p}'_{d+1} = \gamma$. It is easy to check that $\mathrm{H}_1(\boldsymbol{p}') = -(1-\gamma)\log_2(1-\gamma) - \gamma\log_2\gamma + (1-\gamma)\mathrm{H}_1(\boldsymbol{p})$. Setting $\gamma = \frac{1}{1+2^{\mathrm{H}_1(\boldsymbol{p})}}$ we get

$$\begin{aligned}
\mathrm{H}_1(\boldsymbol{p}') - \mathrm{H}_1(\boldsymbol{p}) &= -(1-\gamma)\log_2(1-\gamma) - \gamma\log_2\gamma - \gamma\mathrm{H}_1(\boldsymbol{p}) \\
&\quad - (1-\gamma)\log_2(1-\gamma) - \gamma\log_2\left(2^{\mathrm{H}_1(\boldsymbol{p})}\gamma\right) \\
&= \log_2\frac{2^{\mathrm{H}_1(\boldsymbol{p})}+1}{2^{\mathrm{H}_1(\boldsymbol{p})}} \\
&\geqslant \log_2\frac{d+1}{d} \\
&\geqslant (1-b)\log_2\frac{d}{d-1}
\end{aligned}$$

where we use $b \geqslant \frac{1}{d}$ in the last line. Since $\mathrm{H}_1(\boldsymbol{p}') - \mathrm{H}_1(\boldsymbol{p}) = 0$ for $\gamma = 0$, by the continuity we conclude that there exists $\gamma = \gamma_b$ such that $\boldsymbol{p}'$ satisfies

$$(1-b)\log_2\frac{d+1}{d} = \mathrm{H}_1(\boldsymbol{p}') - \mathrm{H}_1(\boldsymbol{p}).$$

Now we see that the same b solves Equation (7) with the dimension $d' = d+1$ and the constraint $k' = \mathrm{H}_1(\boldsymbol{p}')$. By $\mathrm{H}_1(\boldsymbol{p}') - \mathrm{H}_1(\boldsymbol{p}) \geqslant \log_2\frac{d+1}{d}$ we conclude that $\Delta' = \log_2(d+1) - \mathrm{H}_1(\boldsymbol{p}') \leqslant \log_2 d - \mathrm{H}_1(\boldsymbol{p}) = \Delta$ so the entropy gap is even smaller. After a finite number of step, we end with $\Delta' \leqslant \Delta$, the same b and $d'\Delta' \geqslant 13$. Then by the first case we get that the squared distance is at most $O(\Delta'^2) = O(\Delta^2)$. □

4 Applications

4.1 Negative Results

Corollary 3 (Shannon Entropy estimators are inefficient as uniformity tests)**.** *Suppose that $n \gg 1$ and $\epsilon > 2^{-0.9n}$. Then there exists a distribution $X \in \{0,1\}^n$ such that $\mathrm{H}_1(X) \geqslant n - \epsilon$ but $\mathrm{SD}(X;U_n) = \Omega(\epsilon/n)$.*

Remark 4. *Note that typically one estimates Shannon Entropy within an additive error $O(1)$. However here, to prove that the distribution is ϵ-close to uniform, one has to estimate the entropy with an error $O(n\epsilon)$, which is much tighter! The best known bounds on the running time for an additive error $O(\epsilon)$ are polynomial in ϵ [AOST14, Hol06][1] . With ϵ secure (meaning small) enough for cryptographic purposes, such a precision is simply not achievable in a reasonable time!*

[1] More precisely they require $\mathrm{poly}(\epsilon^{-1})$ independent samples.

Proof of Corollary 3. Take $d = 2^n$ in Corollary 2 and $\Delta = \epsilon$. Suppose that $\Delta = \Omega(2^{-0.9n})$. we have $\mathrm{d}_2(X; U_n) = \Theta(\Delta n^{-1})$. In the other hand we have the trivial inequality $\mathrm{d}_2(X; U_n) \leqslant 4 \cdot \mathrm{SD}(X; U_n)$ (which is a consequence of well known inequalities for ℓ_p-norms) and the result follows. □

Corollary 4 (Separating Smooth Renyi Entropy and Shannon Entropy)**.** *For any n,δ such that $2^{-n} < \delta < \frac{1}{6}$, there exists a distribution $X \in \{0,1\}^n$ such that $\mathrm{H}(X) \geqslant (1-2\delta)n + \log_2(1-2^{-n})$, $\mathrm{H}_2(X) \leqslant 2\log_2(1/\delta) - 2$ and $\mathrm{H}_2^\epsilon(X) \leqslant \mathrm{H}_2(X) + 1$ for every $\epsilon \leqslant \delta$. For a concrete setting consider $n = 256$ and $\delta = 2^{-10}$. We have $\mathrm{H}(X) > 255$ but $\mathrm{H}_2(X) \leqslant 18$ and $\mathrm{H}_2^\epsilon(X) \leqslant 19$ for every $\epsilon < 2^{-9}$!*

Proof. We use a distribution of the same form as the optimal distribution as for problem (15). Denote $N = 2^n$ and define $\boldsymbol{p}_i = \frac{1-2\delta}{N-1}$ for $i = 2, \ldots, N$, and $\boldsymbol{p}_1 = 2\delta$. It is easy to see that $\mathrm{H}(\boldsymbol{p}) \geqslant (1-2\delta)n + \log_2(1-2^n)$ and $\mathrm{H}_2(\boldsymbol{p}) < \log(1/\delta) - 2$. Consider now arbitrary distribution $\boldsymbol{p}'$ such that $\mathrm{SD}(\boldsymbol{p}; \boldsymbol{p}') \leqslant \epsilon$. We have $\boldsymbol{p}'_i = \boldsymbol{p}_i + \epsilon_i$ where $\sum_i \epsilon_i = 0$ and $\sum_i |\epsilon_i| = 2\epsilon$. Note that

$$\begin{aligned}\sum_{i>1} \boldsymbol{p}'^2_i - \sum_{i>1} \boldsymbol{p}_i^2 &> 2\sum_{i>1} \boldsymbol{p}_i \epsilon_i \\ &> -\frac{2(1-2\delta)\epsilon}{N-1} \\ &= -\frac{2\epsilon}{1-2\delta} \cdot \sum_{i>1} \boldsymbol{p}_i{}^2,\end{aligned}$$

and $\boldsymbol{p}'^2_1 - \boldsymbol{p}_1^2 \geqslant -\delta^2 = -\frac{1}{2}\boldsymbol{p}_1^2$. Thus, for $2\epsilon + \delta < \frac{1}{2}$ it follows that $\sum_{i\geqslant 1} \boldsymbol{p}'^2_i \geqslant \left(1 - \frac{1}{2}\right) \sum_{i \geqslant 1} \boldsymbol{p}_i^2$ and the result follows. □

Corollary 5 (Separating Extractable Entropy and Shannon Entropy)**.** *For any $n \geqslant 1$, $\epsilon \in (0,1)$ and $\delta > 2^{-n}$, there exists a random variable $X \in \{0,1\}^n$ such that $\mathrm{H}(X) \geqslant (1-\epsilon-\delta)n + \log_2(1-2^{-n})$ but $\mathrm{H}^\epsilon_{\mathrm{ext}}(X) \leqslant \log(1/\delta)$. For a concrete setting consider $n = 256$ and $\delta = 2^{-10}$. We have $\mathrm{H}(X) > 255.5$ but $\mathrm{H}^\epsilon_{\mathrm{ext}}(X) \leqslant 10$ for every $\epsilon < 2^{-10}$!*

Proof of Corollary 5. We use a distribution of the same form as the optimal distribution as for problem (15). Fix ϵ, δ (we can assume $\epsilon + \delta < 1$) and denote $N = 2^n$. We define $\boldsymbol{p}_i = \frac{1-\epsilon-\delta}{N-1}$ for $i = 2, \ldots, N$, and $\boldsymbol{p}_1 = \epsilon + \delta$. Note that $\boldsymbol{p}_i < \delta$ for $i \neq 1$. It follows then that $\mathrm{H}^\epsilon_\infty(\boldsymbol{p}) \leqslant \log(1/\epsilon)$. In the other hand, note that $\boldsymbol{p}$ is a convex combination of the distribution uniform over the first $N-1$ points (with the weight $1-\epsilon-\delta$ and a point mass at N (with the weight $\epsilon + \delta$. It follows that Shannon Entropy of $\boldsymbol{p}$ is at least $(1-\epsilon-\delta) \cdot \log_2(N-1)$. □

4.2 Positive Results

Corollary 6 (Collision entropy when the Shannon gap is moderate)**.** *Let $k \leqslant n-1$ and let $X^* \in \{0,1\}^n$ minimizes $\mathrm{H}_2(X)$ subject to $\mathrm{H}(X) \geqslant k$ where $X \in \{0,1\}^n$. Then*

$$2\log_2 n - 2\log_2(n-k) \leqslant \mathrm{H}_2(X^*) \leqslant 2\log_2 n - 2\log_2(n+1-k) + 1. \tag{14}$$

For instance, if $k = 255$ then $15 < \mathrm{H}_2(X^*) < 16$.

Proof of Corollary 6. Let b be the solution to Equation (7) (here we have $d = 2^n$). Since $0 \leqslant H(b) \leqslant 1$ we have $\frac{k}{\log_2(d-1)} \geqslant 1 - b \geqslant \frac{k-1}{\log_2(d-1)}$. We improve the left-hand side inequality a little bit

Claim 1. *We have $1 - \frac{k-1}{\log_2 d} \geqslant b \geqslant 1 - \frac{k}{\log_2 d}$.*

Proof of Claim 1. Since $b \geqslant \frac{1}{d}$ we have $\log_2(d-1) - \log(1-b) \geqslant \log_2 d$ and therefore

$$\begin{aligned} k &= -b\log_2 b - (1-b)\log_2(1-b) + (1-b)\log_2(d-1) \\ &\geqslant -b\log_2 b + (1-b)\log_2 d \end{aligned}$$

from which it follows that $1 - b \leqslant \frac{k}{\log_2 d}$. The left part is already proved. □

The result now easily follows by observing that $\frac{(1-b)^2}{d-1} \geqslant b^2$ holds true for $b \leqslant \frac{-1+\sqrt{d-1}}{d-2} \leqslant \frac{1}{2}$, also for $d = 2$. This is indeed satisfied by Claim 1 and $k \leqslant \log_2 d - 1$. □

5 Conclusion

Our results put in a quantitative form the well-accepted fact that Shannon Entropy does not have good cryptographic properties, unless additional strong assumptions are imposed on the entropy source. The techniques we applied may be of independent interests.

Acknowledgment

References

[AOST14] Jayadev Acharya, Alon Orlitsky, Ananda Theertha Suresh, and Himanshu Tyagi, *The complexity of estimating rényi entropy*, CoRR **abs/1408.1000** (2014).

[BBM95] Charles Bennett, Gilles Brassard, and Ueli M. Maurer, *Generalized privacy amplification*, IEEE Transactions on Information Theory **41** (1995), 1915–1923.

[BDK+11] Boaz Barak, Yevgeniy Dodis, Hugo Krawczyk, Olivier Pereira, Krzysztof Pietrzak, Francois-Xavier Standaert, and Yu Yu, *Leftover hash lemma, revisited*, Cryptology ePrint Archive, Report 2011/088, 2011, http://eprint.iacr.org/.

[BK12] Elaine B. Barker and John M. Kelsey, *Sp 800-90a. recommendation for random number generation using deterministic random bit generators*, Tech. report, Gaithersburg, MD, United States, 2012.

[BKMS09] Jan Bouda, Jan Krhovjak, Vashek Matyas, and Petr Svenda, *Towards true random number generation in mobile environments*, Identity and Privacy in the Internet Age (Audun Jsang, Torleiv Maseng, and SveinJohan Knapskog, eds.), Lecture Notes in Computer Science, vol. 5838, Springer Berlin Heidelberg, 2009, pp. 179–189 (English).

[BL05] Marco Bucci and Raimondo Luzzi, *Design of testable random bit generators*, Cryptographic Hardware and Embedded Systems CHES 2005 (JosyulaR. Rao and Berk Sunar, eds.), Lecture Notes in Computer Science, vol. 3659, Springer Berlin Heidelberg, 2005, pp. 147–156 (English).

[BST03] Boaz Barak, Ronen Shaltiel, and Eran Tromer, *True random number generators secure in a changing environment*, In Workshop on Cryptographic Hardware and Embedded Systems (CHES, Springer-Verlag, 2003, pp. 166–180.

[Cac97] Christian Cachin, *Smooth entropy and renyi entropy*, Advances in Cryptology - EUROCRYPT '97, Lecture Notes in Computer Science, SpringerVerlag, 1997, pp. 193–208.

[CW79] J. L. Carter and M. N. Wegman, *Universal classes of hash functions*, Journal of Computer and System Sciences **18** (1979), no. 2, 143–154.

[DPR+13] Yevgeniy Dodis, David Pointcheval, Sylvain Ruhault, Damien Vergniaud, and Daniel Wichs, *Security analysis of pseudo-random number generators with input: /dev/random is not robust*, Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security (New York, NY, USA), CCS '13, ACM, 2013, pp. 647–658.

[HH08] Abdolhossein Hoorfar and Mehdi Hassani, *Inequalities on the lambert w function and hyperpower function*, J. Inequal. Pure and Appl. Math **9** (2008), no. 2.

[HILL88] Johan Hstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby, *Pseudo-random generation from one-way functions*, PROC. 20TH STOC, 1988, pp. 12–24.

[Hol06] Thomas Holenstein, *Pseudorandom generators from one-way functions: A simple construction for any hardness*, Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, Lecture Notes in Computer Science, vol. 3876, Springer, 2006, pp. 443–461.

[LPR11] Cédric Lauradoux, Julien Ponge, and Andrea Röck, *Online Entropy Estimation for Non-Binary Sources and Applications on iPhone*, Rapport de recherche, Inria, June 2011.

[Mau92] Ueli Maurer, *A universal statistical test for random bit generators*, Journal of cryptology **5** (1992), 89–105.

[NZ96] Noam Nisan and David Zuckerman, *Randomness is linear in space*, J. Comput. Syst. Sci. **52** (1996), no. 1, 43–52.

[RTS00] Jaikumar Radhakrishnan and Amnon Ta-Shma, *Bounds for dispersers, extractors, and depth-two superconcentrators*, SIAM JOURNAL ON DISCRETE MATHEMATICS **13** (2000), 2000.

[RW04] R. Renner and S. Wolf, *Smooth Renyi entropy and applications*, International Symposium on Information Theory, 2004. ISIT 2004. Proceedings., IEEE, 2004, p. 232.

[RW05] Renato Renner and Stefan Wolf, *Simple and tight bounds for information reconciliation and privacy amplification*, Proceedings of the 11th International Conference on Theory and Application of Cryptology and Information Security (Berlin, Heidelberg), ASIACRYPT'05, Springer-Verlag, 2005, pp. 199–216.

[Sha48] C. E. Shannon, *A mathematical theory of communication*, Bell system technical journal **27** (1948).

[VSH11] Jonathan Voris, Nitesh Saxena, and Tzipora Halevi, *Accelerometers and randomness: Perfect together*, Proceedings of the Fourth ACM Conference on Wireless Network Security (New York, NY, USA), WiSec '11, ACM, 2011, pp. 115–126.

.

A Proof of Theorem 1

Proof of Theorem 1. Consider the corresponding optimization problem

$$\begin{aligned} \underset{\boldsymbol{p}\in\mathbb{R}^d}{\text{minimize}} \quad & -\log_2\left(\sum_{i=1}^d \boldsymbol{p}_i^2\right) \\ \text{subject to} \quad & 0<\boldsymbol{p}_i,\ i=1,\dots,d. \\ & \sum_{i=1}^d \boldsymbol{p}_i - 1 = 0 \\ & \sum_{i=1}^d -\boldsymbol{p}_i\log_2\boldsymbol{p}_i = k \end{aligned} \tag{15}$$

The Lagrangian associated to (15) is given by

$$L(\boldsymbol{p},(\lambda_1,\lambda_2)) = -\log_2\left(\sum_{i=1}^d \boldsymbol{p}_i^2\right) - \lambda_1\left(\sum_{i=1}^d \boldsymbol{p}_i - 1\right) - \lambda_2\left(-\sum_{i=1}^d \boldsymbol{p}_i\log_2\boldsymbol{p}_i - k\right) \tag{16}$$

Claim 2. *The first and second derivative of the Lagrangian (16) are given by*

$$\frac{\partial L}{\partial \boldsymbol{p}_i} = -2\log_2\mathrm{e}\cdot\frac{\boldsymbol{p}_i}{\boldsymbol{p}^2} - \lambda_1 + \lambda_2\log_2\mathrm{e} + \lambda_2\log_2\boldsymbol{p}_i \tag{17}$$

$$\frac{\partial^2 L}{\partial \boldsymbol{p}_i\boldsymbol{p}_j} = 4\log_2\mathrm{e}\cdot\frac{\boldsymbol{p}_i\boldsymbol{p}_j}{(\boldsymbol{p}^2)^2} + [i=j]\cdot\left(-\frac{2\log_2\mathrm{e}}{\boldsymbol{p}^2} + \frac{\lambda_2\log_2\mathrm{e}}{\boldsymbol{p}_i}\right) \tag{18}$$

Claim 3. *Let $\boldsymbol{p}^*$ be a non-uniform optimal point to 15. Then it satisfies $\boldsymbol{p}_i^*\in\{a,b\}$ for every i, where a,b are some constant such that*

$$-\frac{2\log_2\mathrm{e}}{\boldsymbol{p}^{*2}} + \frac{\lambda_2\log_2\mathrm{e}}{a} > 0 > -\frac{2\log_2\mathrm{e}}{\boldsymbol{p}^{*2}} + \frac{\lambda_2\log_2\mathrm{e}}{b} \tag{19}$$

Proof of Claim 3. At the optimal point $\boldsymbol{p}$ we have $\frac{\partial L}{\partial \boldsymbol{p}_i} = 0$ which means

$$-2\log_2\mathrm{e}\cdot\frac{\boldsymbol{p}_i}{\boldsymbol{p}^2} - \lambda_1 + \lambda_2\log_2\mathrm{e} + \lambda_2\log_2\boldsymbol{p}_i = 0, \quad i=1,\dots,d. \tag{20}$$

Think of $\boldsymbol{p}^2$ as a constant, for a moment. Then the left-hand side of Equation (20) is of the form $-c_1\boldsymbol{p}_i + c_2\log_2\boldsymbol{p}_i + c_3$ with some positive constant c_1 and real constants c_2, c_3. Since the derivative of this function equals $-c_1 + \frac{c_2}{\boldsymbol{p}_i}$, the left-hand side is either decreasing (when $c_2\leqslant 0$) or concave (when $c_2>0$). For the non-uniform solution the latter must be true (because otherwise $\boldsymbol{p}_i$ for $i=1,\dots,d$ are equal). Hence the equation Equation (20) has at most two solutions $\{a,b\}$, where $a<b$ and both are not dependent on i. Moreover, its left-hand side has the maximum between a and b, thus we must have $-c_1+\frac{c_2}{a} > 0 > -c_1+\frac{c_2}{b}$. Expressing this in terms of λ_1,λ_2 we get Equation (19). □

Claim 4. *Let $\boldsymbol{p}^*$ and a,b be as in Claim 3. Then $\boldsymbol{p}_i = a$ for all but one index i.*

Proof of Claim 4. The tangent space of the constraints $\sum_{i=1}^{d} \boldsymbol{p}_i - 1 = 0$ and $-\sum_{i=1}^{d} \boldsymbol{p}_i \log_2 \boldsymbol{p}_i - k = 0$ at the point $\boldsymbol{p}$ is the set of all vectors $h \in \mathbb{R}^d$ satisfying the following conditions

$$\begin{aligned} \textstyle\sum_{i=1}^{d} h_i &= 0 \\ \textstyle\sum_{i=1}^{d} -(\log_2 \boldsymbol{p}_i + \log_2 \mathrm{e}) h_i &= 0 \end{aligned} \tag{21}$$

Intuitively, the tangent space includes all infinitesimally small movements that are consistent with the constraints. Let $D^2 L = \left(\frac{\partial^2 L}{\partial \boldsymbol{p}_i \boldsymbol{p}_j}\right)_{i,j}$ be the second derivative of L. It is well known that the necessary second order condition for the minimizer $\boldsymbol{p}$ is $h^T (D^2) L h \geqslant 0$ for all vectors in the tangent space (21). We have

$$h^T \cdot (D^2 L) \cdot h = 4 \log_2 \mathrm{e} \cdot \frac{\left(\sum_{i=}^{d} \boldsymbol{p}_i h_i\right)^2}{(\boldsymbol{p}^2)^2} + \sum_{i=1}^{d} \left(-\frac{2 \log_2 \mathrm{e}}{\boldsymbol{p}^2} + \frac{\lambda_2 \log_2 \mathrm{e}}{\boldsymbol{p}_i} \right) h_i^2.$$

Now, if the are two different indexes i_1, i_2 such that $\boldsymbol{p}^*_{i_1} = \boldsymbol{p}^*_{i_2} = b$, we can define $h_{i_1} = -\delta$, $h_{i_2} = \delta$ and $h_i = 0$ for $i \notin \{i_1, i_2\}$. Then we get

$$h^T \cdot (D^2 L) \cdot h = 2 \left(-\frac{2 \log_2 \mathrm{e}}{\boldsymbol{p}^2} + \frac{\lambda_2 \log_2 \mathrm{e}}{b} \right) \delta^2$$

which is negative according to Equation (19). Thus we have reached a contradiction. □

Finally, taking into account the case of possibly uniform $\boldsymbol{p}^*$ and combining it with the last claim we get

Claim 5. *The optimal point $\boldsymbol{p}^*$ satisfies $\boldsymbol{p}^*_{i_0} = b$ and $\boldsymbol{p}^*_i = \frac{1-b}{d-1}$ for $i \neq i_0$, for some $b \geqslant \frac{1}{d}$. Then we have $\mathrm{H}(\boldsymbol{p}^*) = \mathrm{H}(b) + (1-b)\log_2(d-1)$ and $\mathrm{H}_2(\boldsymbol{p}^*) = -\log_2 \left(b^2 + \frac{(1-b)^2}{d-1} \right)$.*

It remains to take a closer look at Equation (7). It defines b as an *implicit function* of k and d. Its uniqueness is a consequence of the following claim

Claim 6. *The function $f(b) = \mathrm{H}(b) + (1-b)\log_2(d-1)$ is strictly decreasing and concave for $b \geqslant \frac{1}{d}$.*

Proof of Claim 6. The derivative equals $\frac{\partial f}{\partial b} = -\log_2 \frac{b}{1-b} - \log_2(d-1)$ and hence, for $\frac{1}{d} < b < 1$, is at most $-\log_2 \frac{\frac{1}{d}}{1-\frac{1}{d}} - \log_2(d-1) = 0$. The second derivative is $\frac{\partial^2 f}{\partial b^2} = -\frac{\log_2 \mathrm{e}}{b(1-b)}$. Thus, the claim follows. □

The statement follows now by Claim 5 and Claim 6. □

B Proof of Lemma 2

Proof. Let $\Delta = \log_2 d - k$ be the gap in the Shannon Entropy. Note that from Equation (7) and the inequality $-2 \leqslant d(\log_2(d-1) - \log_2 d) \leqslant -\log_2 \mathrm{e}$ it follows that

$$-b \log_2 b - (1-b)\log_2(1-b) - b \log_2 d = -\Delta + C_1(d) \cdot d^{-1}$$

where $\log_2 \mathrm{e} \leqslant C_1 \leqslant 2$. Note that $f\left(\frac{1}{2}\right) = -1 + \frac{1}{2}\log_2(d-1) < \log_2 d - 1$. Since $\Delta \leqslant 1$ implies $f(b) \geqslant \log_2 d - 1$, by Claim 6 we conclude that $b < \frac{1}{2}$. Next, observe that $1 \leqslant \frac{-(1-b)\log_2(1-b)}{b} \leqslant \log_2 \mathrm{e}$,

for $0 < b < \frac{1}{2}$. This means that $-(1-b)\log_2(1-b) = -b\log_2 C_2(d)$ where $\frac{1}{\mathrm{e}} \leqslant C_2(d) \leqslant \frac{1}{2}$. Now we have

$$-b\log_2(C_2(d)\cdot d\cdot b) = -\Delta + C_1(d)\cdot d^{-1}.$$

Let $y = C_2(d)\cdot d\cdot b$. Our equation is equivalent to $y\ln y = C_3(d)\cdot d\cdot\Delta - C_1(d)C_3(d)$. where $C_3 = C_2/\log_2\mathrm{e}$. Using the Lambert-$W$ function, which is defined as $W(x)\cdot\mathrm{e}^{W(x)} = x$, we can solve this equations as

$$b = \frac{\mathrm{e}^{W(C_3(d)d\Delta - C_3(d)C_1(d))}}{C_2(d)d}. \tag{22}$$

For $x \geqslant \mathrm{e}$ we have the well-known approximation for the Lambert W function [HH08]

$$\ln x - \ln\ln x < W(x) \leqslant \ln x - \ln\ln x + \ln(1+\mathrm{e}^{-1}). \tag{23}$$

Provided that $C_3(d)d\Delta - C_3(d)C_1(d) \geqslant 1$, which is satisfied if $d\Delta \geqslant 6$, we obtain

$$b = \frac{C_3(d)d\Delta - C_3(d)C_1(d)}{C_3(d)d\cdot\log_2\left(C_3(d)d\Delta - C_3(d)C_1(d)\right)}\cdot C_4(d) \tag{24}$$

where $1 \leqslant C_4(d) \leqslant 1+\mathrm{e}^{-1}$. Since the function $x \to \frac{x}{\log_2 x}$ is increasing for $x \geqslant \mathrm{e}$ and since for $d\Delta \geqslant 13$ we have $C_3(d)d\Delta - C_3(d)C_1(d) \geqslant \mathrm{e}$, from Equation (24) we get

$$b \leqslant \frac{C_3(d)d\Delta}{C_3(d)d\cdot\log_2\left(C_3(d)d\Delta\right)}\cdot C_4(d) = \frac{C_4(d)\Delta}{\log_2\left(C_3(d)d\Delta\right)} \tag{25}$$

from which the right part of Equation (8) follows by the inequalities on C_3 and C_4. For the lower bound, note that for $d\Delta \geqslant 13$ we have $C_3(d)d\Delta - C_3(d)C_1(d) \geqslant C_3(d)d\Delta\cdot\frac{11}{13}$ because it reduces to $C_1(d) \leqslant 2$, and that $C_3(d)d\Delta\cdot\frac{11}{13} \geqslant 13\cdot\frac{1}{\mathrm{e}\log_2\mathrm{e}}\cdot\frac{11}{13} > \mathrm{e}$. Therefore, by Equation (24) and the mononicity of $\frac{x}{\log_2 x}$ we get

$$b \geqslant \frac{\frac{11}{13}C_3(d)d\Delta}{C_3(d)d\cdot\log_2\left(\frac{11}{13}C_3(d)d\Delta\right)}\cdot C_4(d) = \frac{\frac{11}{13}C_4(d)\Delta}{\log_2\left(\frac{11}{13}C_3(d)d\Delta\right)}, \tag{26}$$

from which the left part of Equation (8) follows by the inequalities on C_3 and C_4. □