Shannon Entropy Versus Renyi Entropy from a Cryptographic Viewpoint

Maciej Skórski(✉)

Cryptology and Data Security Group, University of Warsaw, Warsaw, Poland
maciej.skorski@mimuw.edu.pl

Abstract. We provide a new inequality that links two important 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. While in practice it is easier to evaluate Shannon entropy than other entropy notions, it is well known in folklore that it does not provide a good estimate of randomness quality from a cryptographic viewpoint, except very special settings. Our results and techniques put this in a quantitative form, allowing us to precisely answer the following questions:

(a) How accurately does Shannon entropy estimate uniformity? Concretely, if the Shannon entropy of an n-bit source X is $n-\epsilon$, where ϵ is a small number, can we conclude that X is close to uniform? This question is motivated by uniformity tests based on entropy estimators, like Maurer's Universal Test.

(b) How much randomness can we extract having high Shannon entropy? That is, if the Shannon entropy of an n-bit source X is $n-O(1)$, how many almost uniform bits can we retrieve, at least? This question is motivated by the folklore upper bound $O(\log(n))$.

(c) Can we use high Shannon entropy for key derivation? More precisely, if we have an n-bit source X of Shannon entropy $n-O(1)$, can we use it as a secure key for some applications, such as square-secure applications? This is motivated by recent improvements in key derivation obtained by Barak et al. (CRYPTO'11) and Dodis et al. (TCC'14), which consider keys with some entropy deficiency.

Our approach involves convex optimization techniques, which yield the shape of the "worst" distribution, 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: Shannon entropy · Renyi entropy · Smooth renyi entropy · Min-entropy · Lambda w function

A full version of this paper is available at https://eprint.iacr.org/2014/967.pdf.

M. Skórski — This work was partly supported by the WELCOME/2010-4/2 grant founded within the framework of the EU Innovative Economy Operational Programme.

J. Groth (Ed.): IMACC 2015, LNCS 9496, pp. 257–274, 2015.
DOI: 10.1007/978-3-319-27239-9_16

1 Introduction

1.1 Entropy Measures

Entropy, as a measure of randomness contained in a probability distribution, is a fundamental concept in information theory and cryptography. There exist many entropy definitions and they are not equally good for all applications. While the most famous (and most liberal) Shannon Entropy [Sha48], which quantifies the encoding length, is extremely useful in information theory, more conservative measures, like min-entropy (which quantifies unpredictability) or collision entropy (which bounds collision probability between samples), are necessary in cryptographic applications, like extracting randomness [NZ96, HILL99, RW05] or key derivation [DY13, BDK+11]. Any misunderstanding about what is a suitable entropy notion may be a serious problem not only of a theoretical concern, because it leads to vulnerabilities due to overestimating security. In fact, when entropy is overestimated, security of real-world applications can be broken [DPR+13]. Standards [BK12, AIS11] recommend to use more conservative entropy metrics in practical designs, but in the other hand Shannon entropy is easier to evaluate [AIS11] (in particular when the distribution of the randomness source is not exactly known) and moreover Shannon entropy estimators have already been relatively well studied and are being used in practice [Mau92, Cor99, BL05, LPR11].

1.2 Motivations and Goals of this Work

The aim of this paper is to provide sharp separation results between Shannon entropy and Renyi entropy (focusing on collision entropy and min-entropy). Under certain conditions, for example when consecutive bits of a given random variable are independent (produced by a memoryless source), they are comparable [RW05, Hol11] (this observation is closely related to a result in information theory known as the Asymptotic Equipartition Property [Cac97]). Such a simplifying assumption is used to argue about provable security of true random number generators [BKMS09, VSH11, LPR11], and may be enforced in certain settings, for example when certifying devices in a laboratory [BL05]. But in general (especially from a theoretical viewpoint) neither min-entropy (being of fundamental importance for general randomness extraction [RW05, Sha11]) nor collision entropy, useful for key derivation [DY13, BDK+11, Shi15], randomness extraction [HILL99], and random number generating [BKMS09, BST03]) *cannot* be well estimated by Shannon entropy. Still, in practice Shannon entropy remains an important tool for testing cryptographic quality of randomness [AIS11]. In this paper we address the natural question

> How bad is Shannon entropy as an estimate of cryptographic quality of randomness?

and answer it in a series of bounds, focusing on three important cryptographic applications, which require entropy estimation: (a) uniformity testing, (b) general randomness extraction and (c) key derivation.

1.3 Our Results and Techniques

Brief Summary. We investigate in 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, identifying the worst case. Our results are mostly negative, in the sense that the gap may be very big, so that even almost full Shannon Entropy does not guarantee that the given distribution is close to uniform or that it may used to derive a secure key. This agrees with folklore. However, to the best of our knowledge, our analysis for the first time provides a comprehensive and detailed study of this problem, establishing tight bounds. Moreover, our techniques may be of independent interests and can be extended to compare Renyi entropy of different orders.

Results and Corollaries. *Bounding Renyi Entropy by Shannon Entropy.* Being 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 $H_1(X)$ of an n-bit random variable X is at least k. What is the best lower bound on the collision entropy $H_2(X)$?

We give a complete answer to this question in Sect. 3.1. It is briefly summarized in Table 1 below.

Table 1. Minimal collision entropy given Shannon entropy constraints.

Domain of X	$H_1(X)$	Region	Max. ℓ_2-distance to uniform	Min. value of $H_2(X)$
$\{0,1\}^n$	$n-\Delta$	$2^n\Delta \geqslant 13$	$\Theta\left(\frac{\Delta}{\log(2^n\Delta)}\right)$	$n-\log_2\left(1+\Theta\left(2^n\Delta^2\log^{-2}(2^n\Delta)\right)\right)$
		$2^n\Delta \leqslant 13$	$O(\Delta)$	$n-\log_2\left(1+O\left(2^n\Delta^2\right)\right)$

The Shape of the Worst-case Distribution. Interestingly, the description of the "worst" distribution X is pretty simple: it is 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]. However, as far as we know, our paper is the first one which provides a full proof that this shape is really best possible.

Infeasibility of Uniformity Tests Based on Entropy Estimators. If an n-bit random variable X satisfies $H_1(X) = n$ then it must be uniform. It might be tempting to think that a very small entropy gap $\Delta = n - H_1(X)$ (when entropy is very "condensed") implies closeness to the uniform distribution. Clearly, this is a necessary condition. For example, standards for random number generating [AIS11] require the Shannon entropy of raw bits to be at least 0.997 per bit on average.

> **Q**: Suppose that the Shannon Entropy $H_1(X)$ of an n-bit random variable X is at least $n - \Delta$, where $\Delta \approx 0$. What is the best upper bound on the distance between X and the uniform distribution U_n?

There are popular statistical randomness tests [Mau92, Cor99] which are based on the fact that very small Δ is necessary to a very small statistical distance. Theoretically, they can detect any deviation at any confidence level. In this paper we quantify what is well known in folklore, namely that this approach *cannot be provable secure and efficient* at the same time. Based on the results summarized in Table 1, we prove that for the statistical distance (the ℓ_1 distance) the gap Δ can be as small as ϵ but still the source is ϵ/n-far from the uniform distribution. Putting this statement around, to guarantee ϵ-closeness we need to estimate the entropy up to a tiny margin $n\epsilon$. This shows that an application of entropy estimators to test sequences of truly random bits may be problematic, because estimating entropy within such a small margin is computationally inefficient. Having said this, we stress that entropy estimators like Maurer's Universal Test [Mau92] are powerful tools capable of discovering most of defects which appear within a broader margin of error.

Large Gap Between Shannon and Smooth Collision Entropy. Many constructions in cryptography require min-entropy. However, the weaker notion of collision entropy found also many applications, especially for problems when one deals with imperfect randomness. The collision entropy of a distribution X constitutes a lower bound on the number of extractable almost-uniform bits, according to the Leftover Hash Lemma [HILL99, RW05]. Moreover, the recent improvements in key derivation [DY13, BDK+11] show that for some applications we can use high collision entropy to generate secure keys, wasting much less entropy comparing to extractors-based techniques (see Sect. 2.5). For example, consider the one-time MAC with a 160-bit key over $GF(2^{80})$, where the key is written as (a, b) and the tag for a message x is $ax + b$. The security is $\epsilon = 2^{-80}$ when the key is uniform [DY13]. We also know that it is $\epsilon = 2^{-70}$-secure when the key has $150 = 160 - 10$ bits of collision entropy. Suppose that a Shannon entropy estimator indicates 159 bits of entropy. Is our scheme secure? This discussion motivates the following question

> **Q**: Suppose that the Shannon Entropy $H_1(X)$ of a random variable $X \in \{0,1\}^n$ is at least $n - \Delta$ where $\Delta \leqslant 1$. What is the best lower bound on $H_2(X)$? Does it help if we consider only $H_2(X')$ 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 weak security requirements, 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. This is just an illustrative example, we provide a more general analysis in Corollary 4 in Sect. 4.

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 [RW05].

> **Q**: Suppose that the Shannon Entropy $H_1(X)$ 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! This is just an illustrative example, we provide a more precise and general statement. 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 the proof of the results in Table 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 small amount of Renyi Entropy in every distribution of high Shannon Entropy.

> **Q**: Suppose that the Shannon Entropy of an n-bit random variable X is at least $n - \Delta$ where $\Delta \geqslant 1$. Does X have some non-trivial amount of collision entropy?

This question may be relevant in settings, when one would like to check whether some (not really big though) collision entropy is present in the source. For example, there are necessary conditions on security of message authentication codes in terms of collision entropy [Shi15]. 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. Again, besides this numerical example we provide detailed and general bounds.

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.

1.4 Organization of the Paper

We start with necessary definitions and explanations of basic concepts in Sect. 2. Our main result is discussed in Sect. 3. Further applications are given in Sect. 4. We end with the conclusion in Sect. 5. The proofs of main results, which are technical and complicated a bit, appear in Sect. 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 probability mass function of a random variable X is denoted by P_X.

2.2 Quantifying Closenes of Distributions

The closeness of two distributions X, Y over the same domain Ω 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 $\mathrm{SD}(X;Y) = \frac{1}{2}\mathrm{d}_1(P_X;P_Y) = \frac{1}{2}\sum_x |\Pr[X=x] - \Pr[Y=x]|$. In this paper we use also the ℓ_2-distance between probability distributions, defined as $\mathrm{d}_2(P_X;P_Y) = \sqrt{\sum_x (\Pr[X=x] - \Pr[Y=x])^2}$. These two ℓ_p distances are related by $d_2(\cdot) < d_1(\cdot) \leqslant \sqrt{|\Omega|} \cdot d_2(\cdot)$. In information theory the closeness of two distributions is often measures using so called *divergences*. The Kullback-Leibler divergence between X and Y is defined as $\mathrm{KL}(X \parallel Y) = -\sum_x P_X(x) \log \frac{P_X(x)}{P_Y(x)}$, and the Renyi divergence of order 2 equals $D_2(X \parallel Y) = \sum_x \frac{(P_X(x)-P_Y(x))^2}{P_X(x)}$. We have $D_2(X \parallel U_S) = H_2(U_S) - H_2(X) = \log_2(|S|\mathrm{CP}(X))$.

For convenience we define also the collision probability of X as the probability that two independent copies of X collide: $\mathrm{CP}(X) = \sum_x \Pr[X=x]^2$.

2.3 Entropy Definitions

Below we define the three key entropy measures, already mentioned in the introduction. It is worth 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* $\mathrm{H}(X) = H_1(X)$*, the collision entropy (or Renyi entropy)* $H_2(X)$*, and the Min-Entropy* $\mathrm{H}_\infty(X)$ *of a distribution* X *are defined as follows*

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

$$H_2(X) = -\log\left(\sum_x \Pr[X=x]^2\right) \tag{2}$$

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

Remark 1 (Comparing Different Entropies). It is easy to see that we have

$$\mathrm{H}(X) \geqslant H_2(X) \geqslant \mathrm{H}_\infty(X),$$

with the equality if and only if X is uniform.

2.4 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 of Smooth Entropy. The smoothing technique is typically used to *increase entropy* by cutting off big but rare "peaks" in a probability distribution, that is point masses relatively heavy comparing to others. 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 all n-bit sequences the most likely ones are those with 1 in almost all places. In particular $\mathrm{H}_\infty(X) = -n \log p$. However, for most of the 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 [RW04] revisited the smoothing technique in entropy framework and came up with new applications.

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 $U_{\mathcal{U}}$.

2.5 Randomness Extraction and Key Derivation

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 with all 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*

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

A very simple, efficient and optimal (with respect 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 random m bits, let H be chosen from $\mathcal{H}$ at random and let X be an n-bit variable. If $H_2(X)\geqslant k$, 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 $H^\epsilon_\infty(X)\geqslant H^\epsilon_{\mathrm{ext}}(X)\geqslant H_2^{\epsilon/2}(X)-2\log(1/\epsilon)-1$.*

Note that to extract k bits ϵ-close to uniform we need to invest $k+2\log(1/\epsilon)$ bits of (collision) entropy; the loss of $2\log(1/\epsilon)$ bits here is optimal [RTS00]. While 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]. Note that the LHL uses only collision entropy, weaker than min-entropy.

To get an illustrative example, note that deriving a key which is ϵ-close to uniform with $\epsilon=2^{-80}$ requires losing $L=2\log(1/\epsilon)=160$ bits of entropy. Sometimes we can't afford to lose so much. In special cases, in particular for so called *square-friendly applications* [BDK+11, DY13] we can get an improvement over Corollary 1. In particular, for these applications (which include message authentication codes or digital signatures), we can apply X of collision entropy $k<m$, still achieving some non-trivial security.

Theorem 1 (Beating the $2\log(1/\epsilon)$ Entropy Loss for Some Applications. *[BDK+11]*). *Let P be an ϵ-secure and σ-square-secure application (when keyed with U_m). Let $\mathcal{H}$ be a universal class of functions from n to random m bits, let H be chosen from $\mathcal{H}$ at random. Then for any X of length n and collision-entropy k, the application P keyed with $H(X)$ given H is ϵ'-secure when $\epsilon'\leqslant\epsilon+\sqrt{\sigma}\cdot 2^{\frac{m-k}{2}}$.*

In particular, when $\sigma=\epsilon$ we get around of the RT-bound, achieving $\epsilon'\approx 2\epsilon$ with only $k=m+\log(1/\epsilon)$. This way we save $\log(1/\epsilon)\approx 80$ bits.

3 Main Result

In this section we calculate what is the minimal collision entropy in a distribution having a certain amount of Shannon entropy. First, by means of convex

optimization, we show in Sect. 3.1 that the uniform distribution with one extra heavy mass is the "worst" shape. Next, using some facts about Lambert W function, in Sect. 3.2 we solve the corresponding implicit equation and derive a closed-form answer.

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

$$H_2(X) \geqslant -\log_2\left(b^2 + \frac{(1-b)^2}{d-1}\right) \tag{6}$$

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 Eq. (6) is best possible.

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

The proof of Theorem 2 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 Eq. (7). We will substitute it to Eq. (6) in order to obtain a closed-form expression.

Lemma 2 (The solution for Moderate Gaps). *Let b be the solution to Eq. (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 Eq. (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 inspection of the proof shows that the numerical constants in Lemma 2 can be made sharper, if needed. Under the mild assumption that $\Delta^{-1} = 2^{o(\log_2 d)}$ 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 can make 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 $H_2(X)$ 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 $H_2(X^*) = -\log_2\left(\frac{1}{d} + \mathrm{d}_2(X^*;U)^2\right)$.

Proof (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

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

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 Eqs. (12) and (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 $H_1(\boldsymbol{p}') = -(1-\gamma)\log_2(1-\gamma) - \gamma \log_2 \gamma + (1-\gamma)H_1(\boldsymbol{p})$. Setting $\gamma = \frac{1}{1+2^{H_1(\boldsymbol{p})}}$ we get

$$\begin{aligned} H_1(\boldsymbol{p}') - H_1(\boldsymbol{p}) &= -(1-\gamma)\log_2(1-\gamma) - \gamma \log_2 \gamma - \gamma H_1(\boldsymbol{p}) \\ &\quad - (1-\gamma)\log_2(1-\gamma) - \gamma \log_2 \left(2^{H_1(\boldsymbol{p})}\gamma\right) \\ &= \log_2 \frac{2^{H_1(\boldsymbol{p})}+1}{2^{H_1(\boldsymbol{p})}} \\ &\geqslant \log_2 \frac{d+1}{d} \\ &\geqslant (1-b)\log_2 \frac{d}{d-1} \end{aligned}$$

where the first inequality follows by $H_1(\boldsymbol{p}) \leqslant \log_2 d$, and the second inequality follows because $b \geqslant \frac{1}{d}$ implies that is suffices to prove $\log_2 \frac{d+1}{d} \geqslant \left(1 - \frac{1}{d}\right)\log_2 \frac{d}{d-1}$ or equivalently that $d \log_2 \frac{d+1}{d} \geqslant (d-1)\log_2 \frac{d}{d-1}$; this is true because the map $u \to u \log_2(1+u^{-1})$ is increasing in u for $u > 0$ (we skip an easy argument, which simply checks the derivative). Since $H_1(\boldsymbol{p}') - H_1(\boldsymbol{p}) = 0$ for $\gamma = 0$ and since $H_1(\boldsymbol{p}') - H_1(\boldsymbol{p}) \geqslant (1-b)\log_2 \frac{d}{d-1} > (1-b)\log_2 \frac{d+1}{d}$ for $\frac{1}{1+2^{H_1(\boldsymbol{p})}}$ for $\geqslant (1-b)\log_2 \frac{d}{d-1}$, by continuity we conclude that there exists $\gamma = \gamma_b$, between 0 and $\frac{1}{1+2^{H_1(\boldsymbol{p})}}$, such that $\boldsymbol{p}'$ satisfies

$$(1-b)\log_2 \frac{d+1}{d} = H_1(\boldsymbol{p}') - H_1(\boldsymbol{p}).$$

Adding this Eq. (7) by sides, we conclude that also b solves 7 with the dimension d replaced by $d' = d+1$ and the constraint k replaced by $k' = H_1(\boldsymbol{p}')$. By $H_1(\boldsymbol{p}') - H_1(\boldsymbol{p}) \geqslant \log_2 \frac{d+1}{d}$ we conclude that $\Delta' = \log_2(d+1) - H_1(\boldsymbol{p}') \leqslant \log_2 d - H_1(\boldsymbol{p}) = \Delta$ 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

The first result we provide is motivated by uniformness testing based on Shannon entropy. We hope that n-bit distribution with entropy $n - \Delta$ where $\Delta \approx 0$, that

is with an extremely small entropy defficency, is close to uniform. We show that for this to be true, Δ has to be negligible.

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 $H_1(X) \geqslant n - \epsilon$ 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 within reasonable time.

Proof (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 standard facts about 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 $\mathrm{H}(X) \geqslant (1-2\delta)n + \log_2(1-2^{-n})$, $H_2(X) \leqslant 2\log_2(1/\delta) - 2$ and $\mathrm{H}^{\epsilon}_2(X) \leqslant H_2(X) + 1$ 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 $\mathrm{H}^{\epsilon}_2(X) \leqslant 19$ 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 $H_2(\boldsymbol{p}) < \log(1/\delta) - 2$. 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 $\epsilon < 2^{-10}$!*

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

Proof (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 $\mathrm{H}_\infty^\epsilon(\boldsymbol{p}) \leqslant \log(1/\epsilon)$. 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

Now we address the question what happens when $\Delta > 1$. This is motivated by settings where keys with entropy deficiency can be applied (cf. Theorem 1 and related references).

Corollary 6 (Collision Entropy When the Shannon Gap is Moderate). *Let $k \leqslant n-1$ and let $X^* \in \{0,1\}^n$ minimizes $H_2(X)$ subject to $\mathrm{H}(X) \geqslant k$ where $X \in \{0,1\}^n$. Then*

$$2\log_2 n - 2\log_2(n-k) \leqslant H_2(X^*) \leqslant 2\log_2 n - 2\log_2(n+1-k) + 1. \tag{14}$$

For instance, if $k = 255$ then $15 < H_2(X^*) < 16$.

Proof (Proof of Corollary 6). Let b be the solution to Eq. (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. We have $1 - \frac{k-1}{\log_2 d} \geqslant b \geqslant 1 - \frac{k}{\log_2 d}$.

Proof (Proof of Sect. 4.2). 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 Sect. 4.2 and $k \leqslant \log_2 d - 1$.

4.3 Bounds in Terms of the Renyi Divergence

Our Corollary 2 gives a bound on the ℓ_2-distance between X and U. Note that

$$\mathrm{d}_2(X;U)^2 = \mathrm{CP}(X) - d^{-1} = d^{-1}\left(d\mathrm{CP}(X) - 1\right) = d^{-1}\left(2^{D_2(X\|U)} - 1\right)$$

and thus our bounds can be expressed in terms of the Renyi divergence D_2. Since we find the distribution X with possibly minimal entropy, this gives an *upper bound* on the divergence in terms the Shannon entropy.

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. The author thanks anonymous reviewers for their valuable comments.

A Proof of Theorem 2

Proof (Proof of Theorem 2). Consider the corresponding optimization problem

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

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. 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. 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 (Proof of Appendix A). At the optimal point $\boldsymbol{p}$ we have $\frac{\partial L}{\partial \boldsymbol{p}_i} = 0$ which means

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

Think of $\boldsymbol{p}^2$ as a constant, for a moment. Then the left-hand side of Eq. (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 $i = 1, \ldots, d$ are equal). Hence the Eq. (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 Eq. (19).

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

Proof (Proof of Appendix A). 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 Eq. (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. 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 $H_2(\boldsymbol{p}^*) = -\log_2 \left(b^2 + \frac{(1-b)^2}{d-1} \right)$.

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

Claim. 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 (Proof of Appendix A). 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 Appendices A and B.

B Proof of Lemma 2

Proof. Let $\Delta = \log_2 d - k$ be the gap in the Shannon Entropy. Note that from Eq. (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 Appendix A 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 Eq. (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 Eq. (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 Eq. (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 Eq. (8) follows by the inequalities on C_3 and C_4.

References

[AIS11] A proposal for: Functionality classes for random number generators1, Technical report AIS 30, Bonn, Germany, September 2011. http://tinyurl.com/bkwt2wf

[AOST14] Acharya, J., Orlitsky, A., Suresh, A.T., Tyagi, H.: The complexity of estimating renyi entropy, CoRR abs/1408.1000 (2014)

[BDK+11] Barak, B., Dodis, Y., Krawczyk, H., Pereira, O., Pietrzak, K., Standaert, F.-X., Yu, Y.: Leftover hash lemma, revisited. In: Rogaway, P. (ed.) CRYPTO 2011. LNCS, vol. 6841, pp. 1–20. Springer, Heidelberg (2011)

[BK12] Barker, E.B., Kelsey, J.M.: Sp 800–90a recommendation for random number generation using deterministic random bit generators, Technical report, Gaithersburg, MD, United States (2012)

[BKMS09] Bouda, J., Krhovjak, J., Matyas, V., Svenda, P.: Towards true random number generation in mobile environments. In: Jøsang, A., Maseng, T., Knapskog, S.J. (eds.) NordSec 2009. LNCS, vol. 5838, pp. 179–189. Springer, Heidelberg (2009)

[BL05] Bucci, M., Luzzi, R.: Design of testable random bit generators. In: Rao, J.R., Sunar, B. (eds.) CHES 2005. LNCS, vol. 3659, pp. 147–156. Springer, Heidelberg (2005)

[BST03] Barak, B., Shaltiel, R., Tromer, E.: True random number generators secure in a changing environment. In: Walter, C.D., Koç, Ç.K., Paar, C. (eds.) CHES 2003. LNCS, vol. 2779, pp. 166–180. Springer, Heidelberg (2003)

[Cac97] Cachin, C.: Smooth entropy and rényi entropy. In: Fumy, W. (ed.) EUROCRYPT 1997. LNCS, vol. 1233, pp. 193–208. Springer, Heidelberg (1997)

[Cor99] Coron, J.-S.: On the security of random sources. In: Imai, H., Zheng, Y. (eds.) PKC 1999. LNCS, vol. 1560, p. 29. Springer, Heidelberg (1999)

[CW79] Carter, J.L., Wegman, M.N.: Universal classes of hash functions. J. Comput. Syst. Sci. **18**(2), 143–154 (1979)

[DPR+13] Dodis, Y., Pointcheval, D., Ruhault, S., Vergniaud, D., Wichs, D.: Security analysis of pseudo-random number generators with input: /dev/random is not robust. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security, CCS 2013, pp. 647–658. ACM, New York (2013)

[DY13] Dodis, Y., Yu, Y.: Overcoming weak expectations. In: Sahai, A. (ed.) TCC 2013. LNCS, vol. 7785, pp. 1–22. Springer, Heidelberg (2013)

[HH08] Hoorfar, A., Hassani, M.: Inequalities on the lambert w function and hyperpower function. J. Inequal. Pure Appl. Math. **9**(2), 07–15 (2008)

[HILL88] Hstad, J., Impagliazzo, R., Levin, L.A., Luby, M.: Pseudo-random generation from one-way functions. In: Proceedings of the 20TH STOC, pp. 12–24 (1988)

[HILL99] Hastad, J., Impagliazzo, R., Levin, L.A., Luby, M.: A pseudorandom generator from any one-way function. SIAM J. Comput. **28**(4), 1364–1396 (1999)

[Hol06] Holenstein, T.: Pseudorandom generators from one-way functions: a simple construction for any hardness. In: Halevi, S., Rabin, T. (eds.) TCC 2006. LNCS, vol. 3876, pp. 443–461. Springer, Heidelberg (2006)

[Hol11] Holenstein, T.: On the randomness of repeated experiment

[LPR11] Lauradoux, C., Ponge, J., Röck, A.: Online Entropy Estimation for Non-Binary Sources and Applications on iPhone. Rapport de recherche, Inria (2011)

[Mau92] Maurer, U.: A universal statistical test for random bit generators. J. Cryptology **5**, 89–105 (1992)

[NZ96] Nisan, N., Zuckerman, D.: Randomness is linear in space. J. Comput. Syst. Sci. **52**(1), 43–52 (1996)

[RTS00] Radhakrishnan, J., Ta-Shma, A.: Bounds for dispersers, extractors, and depth-two superconcentrators. SIAM J. Discrete Math. **13**, 2000 (2000)

[RW04] Renner, R., Wolf, S.: Smooth renyi entropy and applications. In: Proceedings of the International Symposium on Information Theory, ISIT 2004, p. 232. IEEE (2004)

[RW05] Renner, R.S., Wolf, S.: Simple and tight bounds for information reconciliation and privacy amplification. In: Roy, B. (ed.) ASIACRYPT 2005. LNCS, vol. 3788, pp. 199–216. Springer, Heidelberg (2005)

[Sha48] Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)

[Sha11] Shaltiel, R.: An introduction to randomness extractors. In: Aceto, L., Henzinger, M., Sgall, J. (eds.) ICALP 2011, Part II. LNCS, vol. 6756, pp. 21–41. Springer, Heidelberg (2011)

[Shi15] Shikata, J.: Design and analysis of information-theoretically secure authentication codes with non-uniformly random keys. IACR Cryptology ePrint Arch. **2015**, 250 (2015)

[VSH11] Voris, J., Saxena, N., Halevi, T.: Accelerometers and randomness: perfect together. In: Proceedings of the Fourth ACM Conference on Wireless Network Security, WiSec 2011, pp. 115–126. ACM, New York (2011)