How much randomness can be extracted from memoryless Shannon entropy sources?

Maciej Skorski
maciej.skorski@mimuw.edu.pl

Cryptology and Data Security Group, University of Warsaw

Abstract. We revisit the classical problem: given a memoryless source having a certain amount of Shannon Entropy, how many random bits can be extracted? This question appears in works studying random number generators built from physical entropy sources.
Some authors use a heuristic estimate obtained from the Asymptotic Equipartition Property, which yields roughly n extractable bits, where n is the total Shannon entropy amount. However the best known precise form gives only $n - O(\sqrt{\log(1/\epsilon)n})$, where ϵ is the distance of the extracted bits from uniform. In this paper we show a matching $n - \Omega(\sqrt{\log(1/\epsilon)n})$ upper bound. Therefore, the loss of $\Theta(\sqrt{\log(1/\epsilon)n})$ bits is necessary. As we show, this theoretical bound is of practical relevance. Namely, applying the imprecise AEP heuristic to a mobile phone accelerometer one might overestimate extractable entropy even by 100%, no matter what the extractor is. Thus, the "AEP extracting heuristic" should not be used without taking the precise error into account.

Keywords: Shannon Entropy, Randomness Extractors, AEP

1 Introduction

1.1 Entropy

Entropy is a measure of randomness in a probability distribution. Since one notion is not capable of covering all possible applications of randomness, we use different entropy definitions depending on a situation. In information theory most widely used is Shannon entropy, which quantifies the encoding length of a given distribution. In turn, cryptographers use the more conservative notion called min-entropy, which quantifies unpredictability. In general, there is a large gap between these two measures: the min-entropy of an n-bit string might be only $O(1)$ whereas its Shannon entropy as big as $\Omega(n)$. However, they are comparable when we have a stream of independent samples. We will return to this case when discussing how to extract from a memoryless Shannon entropy source.

1.2 Randomness Extraction

Some important computer applications, like generating cryptographic keys, coutnermeasures against side-channel attacks or gambling, demand randomness of excellent quality, that is uniform or indistinguishable from uniform bits. Unfortunately, in practice we do not have access to pure randomness. Even best physical sources of randomness produce bits that are slightly biased or slightly correlated. In order to solve this problem, the concept of *randomness extraction* has been developed. In a typical setting one considers *weak sources*, which produce only *high entropy* (rather than uniform) output, and special algorithms called *extractors*, which are capable of converting long sequences of biased bits into shorter but uniform (or very close to) sequences. A device which combines an entropy source and an extractor is called a *true random number generator* [Sun09], in contrast to pseudorandom generators which only expand some initial randomness using deterministic mathematical formulas. The design of a typical TRNG is illustrated in Figure 1 below.

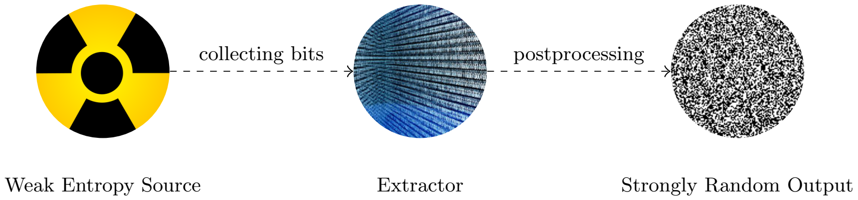

Fig. 1: True Random Number Generators. The scheme illustrates the typical design, where the building blocks are: (a) an entropy source (b) a harvesting mechanism and (c) a posprocessor (extractor).

Entropy Sources In practice, randomness can be gathered based on a physical phenomena (like radiation [hot], photons transmission, thermal noise [BP99], atmospheric noise [ran], jitters) or even from a human-device interaction (like timing events, keystrokes or mouse movements [pgp], shaking accelerators in mobile phones [VSH11] and other).

Extractors Extractors are functions which transform inputs of some required entropy amount into an almost uniform string of known length. To extract from every high-entropy source (that is, to have an extractor of general purpose), one needs to allow extractors to use small amount of auxiliary randomness, which can be "reinvested" as in the case of catalysts in chemistry. Also, one has to accept some small deviation of the output from being uniform (small enough to be acceptable for almost every application) and some entropy loss [RTS00]. Good extractors, simple, provable-secure and widely used in practice, are obtained from universal hash families [CW79]. We refer the reader to [Sha11] for a survey.

1.3 Entropy Estimating

Shannon Entropy vs Min Entropy. For the purpose of cryptographic applications, the right notion of entropy is min-entropy rather than Shannon entropy.

In fact, one can extract randomness only if the given source distribution is close to a distribution of high min-entropy [RW04]. However, in order to avoid overestimating security, we need to compute how much entropy we have; here Shannon entropy is much easier to estimate from samples than min-entropy[1]. In particular, this is the case of estimating entropy of data streams in an *online* manner, at low memory cost. Such estimators are an active research area and find important applications in learning, data mining or network anomaly detection. The problem is that Shannon entropy overestimates crytpographic randomness, being much bigger than min-entropy.

Almost-Equivalence for Stateless Sources. A *stateless source* (called also memoryless) is a source which produces consecutive samples independently. While this is a restriction, it is often assumed by practitioners working on random number generators (cf. [LRSV12,BKMS09,BL05]) and argued to be reasonable for some classes of sources. An important result is obtained from a more general fact called Asymptotic Equipartition Property (AEP). Namelly, for a stateless source the Shannon entropy per bit is close to its min-entropy per bit.

> **A variant of the AEP:** The min entropy per bit in a squence $X_1, \ldots, X_n$ of i.id. samples from X *converges*, when $n \to \infty$, to the Shannon entropy of X. More precisely
>
> $$\frac{H_\infty(X_1,\ldots,X_n)}{n} \overset{\text{in probability}}{\longrightarrow} H(X) \qquad (1)$$

Thus, the AEP is a bridge connecting the heuristic use of Shannon entropy as a measure of extractable randomness (practice) and the provable security (extractors theory). The best known quantitative form appears in [Hol06].

Lemma 1 (Asymptotic Equipartition Property [Hol06]). *Let $X_1, \ldots, X_n$ be i.i.d. samples from a distribution of Shannon entropy k. Then the distribution of $(X_1, \ldots, X_n)$ is ϵ-close to a distribution of min entropy $kn - O\left(\sqrt{\log(1/\epsilon)kn}\right)$.*

We can conclude that, up to an error of $o(n)$, the number of extractable bits (very close to uniform) is at least as big as the Shannon entropy of $(X_1, \ldots, X_n)$. But is it safe to assume (heuristically) that the equality (1) holds in practical parameter regimes?

1.4 Problem statement

We address the problem of finding upper bounds on the extraction rate.

> **Question**: Suppose that we have a source which produces i.i.d samples $X_1, X_2, \ldots$ each of Shannon entropy k. How much almost uniform bits can be extracted from n such samples?

This question is well-motivated as no lower bounds to Lemma 1 are known so far, and because of the popularity of the AEP herustic (1).

[1] One can use, for example, techniques based on Markov chains [Cor99]

1.5 Our Result

The Tight No-Go Result We answer the posted question, showing that the convergence rate in Equation (1) given in Lemma 1 is optimal.

Theorem 1 (An Upper Bound on Bits Extractable from Shannon Sources). *In the above setting, we can extract no more that*

$$N = kn - \Theta(\sqrt{\log(1/\epsilon)kn}) \tag{2}$$

bits which are ϵ-close (in the variation distance) to uniform. This matches the lower bound in [Hol06] (the constant under $\Theta(\cdot)$ depends on the source X).

Corollary 1 (A Significant Entropy Loss in the AEP Heuristic Estimate). *In the above setting, the gap between the Shannon entropy and the number of extractable bits ϵ-close to uniform equals at least $\Theta(\sqrt{\log(1/\epsilon)kn})$. In particular, for the recommended security level ($\epsilon = 2^{-80}$) we obtain the loss of $kn - N \approx \sqrt{80kn}$ bits, no matter what an extractor we use.*

A Practical Example - How Not To Overestimate Security Imagine a mobile phone where the accelerometer is being used as an entropy source. Such a source was studied in [LPR11] and the Shannon entropy rate was estimated to be roughly 0.125 per bit. Since the recommended security level for almost random bits is $\epsilon = 2^{-80}$. According to the heuristic (1) we need roughly $128/0.125 = 1024$ samples to extract a 128-bit key. However taking into account the true error in our Theorem 1 we see that we need at least $n \approx 2214$ bits!

1.6 Organization

In Section 2 we give some basic facts and auxiliary technical results that will be used later. The proof of the main result, that is Theorem 1, is given in Section Section 3. Finally, Section 4 concludes the work.

2 Preliminaries

2.1 Basic Definitions

The most popular way of measuring how two distributions are close is the statistical distance.

Definition 1 (Statistical Distance). *The statistical (or total variation) distance of two distributions X, Y is defined as*

$$\mathrm{SD}\left(X; Y\right) = \sum_{x} \left|\Pr[X = x] - \Pr[Y = x]\right| \tag{3}$$

We also say that X and Y are ϵ-close.

Below we recall the definition of Shannon entropy and min entropy. The logarithms are taken at base 2.

Definition 2 (Shannon Entropy). *The Shannon Entropy of a distribution X equals $H(X) = -\sum_x \Pr[X=x] \log \Pr[X=x]$.*

Definition 3 (Min Entropy). *The min entropy of a distribution X equals $H_\infty(X) = -\max_x \log \Pr[X=x]$.*

Definition 4 (Extractable Entropy, [RW04]). *We say that X has k extractable bits within distance ϵ, denoted $H^{\epsilon}_{\text{ext}}(X) \geqslant k$, if for some randomize function* Ext *we have* $\mathrm{SD}\left(\mathrm{Ext}(X,S); U_k, S\right) \leqslant \epsilon$, *where U_k is a uniform k-bit string and S is an independent uniform string.*

2.2 Technical Facts

Our proof uses the following characterization of "extractable" distributions.

Theorem 2 (An Upper Bound on Extractable Entropy, [RW04]). *If $H^{\epsilon}_{\text{ext}}(X) \geqslant k$ then X is ϵ-close to Y such that $H_\infty(Y) \geqslant k$.*

The second important fact we use is the sharp bound on binomial tails.

Theorem 3 (Tight Binomial Tails [McK]). *Let $B(n,p)$ be a sum of independent Bernoulli trials with success probability p. Then for $\gamma \leqslant \frac{3}{4}q$ we have*

$$\Pr\left[B(n,p) \geqslant pn + \gamma n\right] = Q\left(\sqrt{\frac{n\gamma^2}{pq}}\right) \cdot \psi\left(p,q,n,\gamma\right) \tag{4}$$

with the error term satisfies

$$\psi\left(p,q,n,\gamma\right) = \exp\left(\frac{n\gamma^2}{2pq} - n\mathrm{KL}\left(p+\gamma \parallel p\right) + \frac{1}{2}\log\left(\frac{p+\gamma}{p}\cdot\frac{q}{q-\gamma}\right) + O_{p,q}\left(n^{-\frac{1}{2}}\right)\right) \tag{5}$$

where $\mathrm{KL}\left(a \parallel b\right) = a\log(a/b) + (1-a)\log((1-a)/(1-b)$ *is the Kullback-Leibler divergence, and Q is the complement of the cumulative distribution function of the standard normal distribution.*

3 Proof of Theorem 1

3.1 Characterizing Extractable Entropy

We state the following fact with an explanaition in Figure 2.

Lemma 2 (Lower bound on the extractable entropy). *Let X be a distribution. Then for $\underline{\text{every}}$ distribution Y which is ϵ-close to X, twe have $H_\infty(Y) \leqslant -\log t$ where t satisfies*

$$\sum_x \max(\mathbf{P}_X(x) - t, 0) = \epsilon. \tag{6}$$

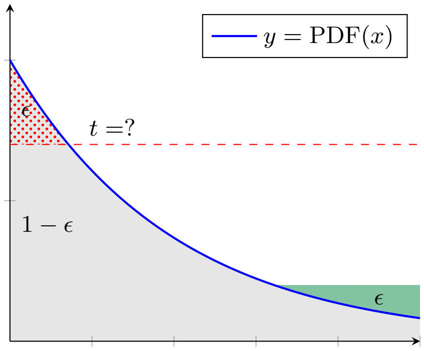

Fig. 2: The Entropy Smoothing Problem. For a given probability density function, we want to cut a total mass of up to ϵ above a possibly highest threshold (in dotted red) and rearrange it (in green), to keep the upper bound smallest possible.

The proof follows easily by observation that the optimal mass rearrangement (which maximizes $H_{\infty}(Y)$) is to decrease probability mass at biggest points. Without losing generality, we assume from now that $X \in \{0,1\}$ where $\Pr[X = 1] = p, q = 1-p$. Define $X^n = (X_1,\ldots,X_n)$. For any $x \in \{0,1\}^n$ we have

$$\Pr[X^n = x] = p^{\|x\|}q^{n-\|x\|}. \tag{7}$$

According to the last lemma and Theorem 2, we have

$$H^{\epsilon}_{\text{ext}}\left(X^n\right) \leqslant -\log t \tag{8}$$

where

$$\sum_x \max\left(\mathbf{P}_{X^n}(x) - t, 0\right) = \epsilon. \tag{9}$$

From now we assume that

$$t = p^{pn+\gamma n}q^{qn-\gamma n}. \tag{10}$$

3.2 Determining the Threshold t

The next key observation is that t is actually small and can be ommitted. That is, we can simply cut the $(1-\epsilon)$-quantile. This is stated in the lemma below.

Lemma 3 (Replacing the threshold by the quantile). *Let $x_0 \in \{0,1\}^n$ be a point such that $\|x_0\| = pn + \gamma n$. Then we have*

$$\sum_{x\colon\ \|x\|\geqslant\|x_0\|} \max\left(\mathbf{P}_{X^n}(x) - \mathbf{P}_{X^n}(x_0)\right) \geqslant \frac{1}{2}\sum_{x\colon\ \|x\|\geqslant\|x_0\|}\mathbf{P}_{X^n}(x) \tag{11}$$

To prove the lemma, note that from Theorem 3 it follows that setting

$$\gamma' = \gamma + n^{-1}\log\left(\frac{p}{q}\right) \tag{12}$$

we obtain

$$\sum_{j \geqslant pn+\gamma' n}\binom{n}{j} \geqslant \frac{3}{4}\cdot\sum_{j \geqslant pn+\gamma n}\binom{n}{j} \tag{13}$$

when γ is sufficiently small comparing to p and q (formally this is justified by calculating the derivative with respect to γ and noticing that it is bigger by at most a factor of $1+\frac{\gamma}{\sqrt{npq}}$). But we also have

$$p^j q^{n-j} \geqslant 2\cdot p^{(p+\gamma)n} q^{(q-\gamma)n} \quad \text{for } j \geqslant \gamma' n \tag{14}$$

Therefore,

$$\begin{aligned}\sum_{j \geqslant pn+\gamma n}\binom{n}{j}p^j q^{n-j} &\geqslant \sum_{j \geqslant pn+\gamma' n}\binom{n}{j}p^j q^{n-j} \\ &\geqslant 2\cdot p^{(p+\gamma)n} q^{(q-\gamma)n}\cdot\sum_{j \geqslant pn+\gamma' n}\binom{n}{j} \\ &\geqslant 2\cdot\frac{3}{4}\cdot p^{(p+\gamma)n} q^{(q-\gamma)n}\cdot\sum_{j \geqslant pn+\gamma n}\binom{n}{j} \qquad (15)\end{aligned}$$

which finishes the proof.

3.3 Putting this all together

Now, by combining Lemma 2, Lemma 3 and the estimate $Q(x) \approx x^{-1}\exp(-x^2/2)$ for $x \gg 0$ we obtain

$$\epsilon \geqslant \exp\left(-n\mathrm{KL}\left(p+\gamma \parallel p\right) - \log\left(\frac{n\gamma^2}{2pq}\right) + O_{p,q}(1)\right) \tag{16}$$

which, because of the Taylor expansion $\mathrm{KL}\left(p+\gamma \parallel p\right) = \frac{\gamma^2}{2pq} + O_{p,q}(\gamma^3)$, gives us

$$\gamma \geqslant \Omega\left(\sqrt{\frac{\log(1/\epsilon)}{pqn}}\right) \tag{17}$$

Setting $\gamma = c\cdot\sqrt{\frac{\log(1/\epsilon)}{pqn}}$, with sufficiently big c, we obtain the claimed result.

4 Conclusion

We show an upper bound on the amount of random bits that can be extracted from a Shannon entropy source. Even in the most favorable case, that is for independent bits, the gap between the Shannon entropy and the amount of randomness that can be extracted is significant. In practical settings the Shannon entropy might be even 2 times bigger than the extractable entropy.

References

BKMS09. Jan Bouda, Jan Krhovjak, Vashek Matyas, and Petr Svenda, *Towards true random number generation in mobile environments*, NordSec 2009, Lecture Notes in Computer Science, vol. 5838, 2009, pp. 179–189.

BL05. Marco Bucci and Raimondo Luzzi, *Design of testable random bit generators*, CHES 2005, vol. 3659, 2005, pp. 147–156 (English).

BP99. Jun Benjamin and Kocher Paul, *The intel random number generator*, 1999.

Cor99. Jean-Sebastien Coron, *On the security of random sources*, 1999.

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.

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

hot. *Hotbits project homepage*, www.fourmilab.ch/hotbits/.

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.

LRSV12. Patrick Lacharme, Andrea Röck, Vincent Strubel, and Marion Videau, *The linux pseudorandom number generator revisited*, Cryptology ePrint Archive, Report 2012/251, 2012, http://eprint.iacr.org/.

McK. Brendan D. McKay, *ON LITTLEWOOD'S ESTIMATE FOR THE BINOMIAL DISTRIBUTION*, Advances in Applied Probability.

pgp. *Pgp project homepage*, http://www.pgpi.org.

ran. *Random.org project homepage*, www.random.org.

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*, ISIT 2004, 2004, p. 232.

Sha11. Ronen Shaltiel, *An introduction to randomness extractors*, ICALP'11, 2011, pp. 21–41.

Sun09. Berk Sunar, *True random number generators for cryptography*, Cryptographic Engineering (etinKaya Ko, ed.), Springer US, 2009, pp. 55–73 (English).

VSH11. Jonathan Voris, Nitesh Saxena, and Tzipora Halevi, *Accelerometers and randomness: Perfect together*, WiSec '11, ACM, 2011, pp. 115–126.