Rose-Hulman Undergraduate Mathematics Journal

Volume 17
Issue 1

Article 7

Randomness Extractors -- An Exposition

Wei Dai
College of Creative Studies, University of California, Santa Barbara

Rose-
Hulman
Undergraduate
Mathematics
Journal

Randomness Extractors – An Exposition

Wei Dai[a]

Volume 17, No. 1, Spring 2016

Sponsored by

Rose-Hulman Institute of Technology
Department of Mathematics
Terre Haute, IN 47803
Email: mathjournal@rose-hulman.edu
http://www.rose-hulman.edu/mathjournal

[a]College of Creative Studies, University of California, Santa Barbara

Rose-Hulman Undergraduate Mathematics Journal
Volume 17, No. 1, Spring 2016

Randomness Extractors – An Exposition

Wei Dai

Abstract. Randomness is crucial to computer science, both in theory and applications. In complexity theory, randomness augments computers to offer more powerful models. In cryptography, randomness is essential for seed generation, where the computational model used is generally probabilistic. However, ideal randomness, which is usually assumed to be available in computer science theory and applications, might not be available to real systems. Randomness extractors are objects that turn "weak" randomness into almost "ideal" randomness (pseudorandomness). In this paper, we will build the framework to work with such objects and present explicit constructions. We will discuss a well-known construction of seeded extractors via universal hashing and present a simple argument to extend such results to two-source extractors.

Acknowledgements: First, I would like to thank my advisors, Prof. Ömer Eğecioğlu, and Prof. Çetin Kaya Koç, for introducing me to the study of random number generators, which led to my study on extractors. I would also like to thank Prof. Ömer Eğecioğlu, Prof. Çetin Kaya Koç, Dr. Elif Saygı and Dr. Zülfükar Saygı for their discussion and suggestions on my writing. I would like to thank the anonymous reviewers for their thoughtful comments. Lastly, I would like to acknowledge that none of this would be possible without my parents, who have supported me financially throughout my undergraduate career.

1 Introduction

Randomness is crucial in computer science, where the usual assumption is access to uniform and independent bits (ideal randomness). However, real systems often fail to offer truly ideal randomness. A weaker property is to require the randomness to be unpredictable. Such a property will be able to capture a larger class of sources such as passwords, biometric data, and system counters. The crucial question is then: How do we derive uniform and independent bits from entropic sources? This paper surveys basic results of randomness extractors, which are objects that map entropic inputs (weak randomness) into almost uniform outputs (pseudorandomness).

Let us fix a finite sample space Ω. The notions of ideal, weak, and pseudorandomness are properties of distributions over Ω. We will discuss these notions informally here. We give formal definitions in Section 2.

Ideal Randomness: By ideal randomness, we simply mean the *uniform* distribution over Ω. In theory, such distributions are very natural to work with. However, if we are working with physical systems, measurements are usually skewed and biased. The theory of randomness extraction tries to bridge between real randomness and ideal randomness.

Weak Randomness: For any distribution, we will define various measures of entropy: Shannon entropy, collision entropy, and minimum entropy. Weak randomness referes to a distribution with at least some measure of entropy. Since many systems assume the availability of "ideal" randomness, we assume that the direct usage of such weak randomness is inadequate.

Pseudorandomness: Roughly speaking, a pseudorandom distribution, even though not uniform, is "good enough" for practical use. If a class of functions, $\mathcal{F}$ (which could represent circuits or programs, for example), cannot distinguish a distribution P from an ideal one, we say that P is pseudorandom with respect to $\mathcal{F}$. This also means that, for all practical purposes, we can use P as input to any function in $\mathcal{F}$ in place of ideal randomness. Mathematically speaking, we will define a (semi)-metric in the space of all distributions with respect to $\mathcal{F}$.

Randomness Extractors: Randomness extractors, or just extractors, are objects that convert weak randomness to pseudorandomness. We will explore the conditions under which such objects exist and show that some simple algebraic functions are good extractors.

We begin with a motivating example of biased coins in Section 1.1 before giving an overview of the rest of the paper in Section 1.2.

1.1 Biased Coins–A Toy Example

Suppose that we are given a coin, and we want to determine whether it is fair or biased. If we are only allowed to toss it once, what is the probability that we can guess right? What if we are allowed k tosses? To make this well defined, we assume that we are given a fair coin (with probabilities of heads and tails each $\frac{1}{2}$), or a biased coin. Let the probability of heads of the biased coin be p for some $p \in [0,1]$. Further, to make the matter easier, assume that the coin tosses are independent of each other. Hence, our game is parameterized by k and p.

Of course, we know k, which is the number of allowed tosses. What about p? We will assume that p follows some probability distribution. For example, it could be 0.9 with probability 1 or uniform over $[0,1]$. Let us first analyze the case $k = 1$ where we are only allowed one toss. After we toss the coin once, we get one outcome, say heads (for simplicity we will use 1 to denote heads and 0 to denote tails). Intuitively, if we do not know anything about the biased coin's probability of heads, p, (in the case where p is uniform in $[0,1]$) we cannot infer anything about whether our coin is fair or biased. However, if we know that $p = 0.9$ let's say, then we get some advantage in guessing that our coin is biased, given that the outcome was heads. On the other hand, if the biased coin is not biased at all ($p = 0.5$), then we cannot possibly distinguish the two coins! What if $p = 0.5 + \epsilon$ and ϵ is very small, say 2^{-128}? In fact, in the theory that we will develop, we say that such a biased coin is ϵ-pseudorandom. Intuitively, the smaller ϵ is, the more difficult it is to distinguish it from a fair coin. What about for $k \geq 1$? Our intuition for $k = 1$ tells us that our ability to guess correctly depends crucially on the distribution of p.

Let us analyze our success probability for general k. If we are allowed k tosses, then our observation is an element of $\{0,1\}^k$, and our output is either 0 for guessing fair or 1 for guessing biased. In other words, guessing from k tosses is a function of the form $\{0,1\}^k \to \{0,1\}$. We will call such a function a k-toss *distinguisher*. Now, for any p and distinguisher D, we can calculate our probability of guessing correctly. For example, suppose $k = 1$, $p = 0.9$ and D is the identity function on $\{0,1\}$. We will let coin 0 be the fair coin and coin 1 be the biased coin. For $b \in \{0,1\}$, let C_b be the outcome of the toss for coin b. Then, our success probability is

$$\begin{aligned}
\Pr[\text{D succeeds}] &= \Pr[D(C_b) = b] \\
&= \Pr[b = 0] \cdot \Pr[D(C_b) = b \mid b = 0] + \Pr[b = 1] \cdot \Pr[D(C_b) = b \mid b = 1] \\
&= \frac{1}{2} \cdot \Pr[C_0 = 0] + \frac{1}{2} \cdot \Pr[C_1 = 1] \\
&= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot 0.9 \\
&= 0.7.
\end{aligned}$$

In general, for any fixed k and p, we can calculate $\Pr[\text{D succeeds} \mid p]$ for any $D : \{0,1\}^k \to \{0,1\}$ as follows. We will let coin 0 be the fair coin and coin 1 be the biased coin. For $b \in \{0,1\}$, let C_b be the outcome of a toss for coin b and $C_b^{(k)}$ be the outcome of k tosses for coin b (recall that tosses for a coin are all independently and identically distributed). Then,

$$\begin{aligned}\Pr[\text{D succeeds}] &= \Pr[D(C_b^{(k)}) = b] \\ &= \frac{1}{2}\Pr[D(C_0^{(k)}) = 0 \mid b = 0] + \frac{1}{2}\Pr[D(C_1^{(k)}) = 1 \mid b = 1] \\ &= \frac{1}{2}\Pr[C_0^{(k)} \in D^{-1}(0)] + \frac{1}{2}\Pr[C_1^{(k)} \in D^{-1}(1)].\end{aligned}$$

Notice that $D^{-1}(0)$ and $D^{-1}(1)$ form a partition of $\{0,1\}^k$, and we can easily compute $D^{-1}(0)$ and $D^{-1}(1)$ given D. Next, under an assumed distribution of the parameter p, we can calculate our distinguisher's estimated probability of success. This is given by

$$\mathbb{E}_p[\Pr[\text{D succeeds} \mid p]],$$

where $\mathbb{E}_p$ denotes the expectation over the distribution of the parameter p. Now, for a fixed distribution of p, we can find a distinguisher D_{max} that maximize our expected probability of success as follows:

$$D_{max} := \underset{D:\{0,1\}^k \to \{0,1\}}{\arg\max} \mathbb{E}_p[\Pr[\text{D succeeds} \mid p]].$$

Another way to look at this problem is to define how "distinguishable" two distributions are. We define the notion of *distinguishability*, given a distinguisher $D : \{0,1\} \to \{0,1\}$. We can evaluate the effectiveness of D in distinguishing a fair coin, C_0, and a biased coin, C_1, as follows

$$\begin{aligned}\Delta_D(C_0, C_1) &:= |\mathbb{E}[D(C_0)] - \mathbb{E}[D(C_1)]| \\ &= |\Pr[D(C_0) = 1] - \Pr[D(C_1) = 1]| \\ &= |\Pr[D(C_0) = 0] - \Pr[D(C_1) = 0]|.\end{aligned}$$

Let $b \in \{0,1\}$ indicate whether the coin is fair or biased. The success probability of a distinguisher D can be related to Δ_D in the following way,

$$\begin{aligned}&\Pr[\text{D succeeds}] \\ &= \Pr[D(C_b) = b] \\ &= \Pr[b = 0]\Pr[D(C_0) = 0] + \Pr[b = 1]\Pr[D(C_1) = 1] \\ &= \frac{1}{2}(\Pr[D(C_0) = 0] - \Pr[D(C_1) = 0] + \Pr[D(C_1) = 0]) + \frac{1}{2}\Pr[D(C_1) = 1] \\ &\leq \frac{1}{2}|\Pr[D(C_0) = 0] - \Pr[D(C_1) = 0]| + \frac{1}{2}(\Pr[D(C_1) = 0] + \frac{1}{2}\Pr[D(C_1) = 1]) \\ &\leq \frac{1}{2}\Delta_D(C_0, C_1) + \frac{1}{2}.\end{aligned}$$

Hence, the distinguisher that maximizes our success probability also maximizes $\Delta_D(C_0, C_1)$. This leads us to the definition of statistical distance: the statistical distance between the fair coin and the biased coin is defined to be

$$\Delta(C_0, C_1) := \max_{D:\{0,1\}\to\{0,1\}} \Delta_D(C_0, C_1).$$

Putting this into the game playing framework that is widely used in cryptography, we are playing the role of an attacker, and the setup and governing rules around what we are allowed to do is acting as what is called the challenger. The challenger picks a fair coin or a biased coin with equal probability and gives the coin to us, the attacker. We perform some tosses and make a guess of whether the coin is fair or biased. We win the game if and only if our guess is correct. We will call this game the biased coin game with parameter k, BCG_k.

In fact, many modern cryptography applications (hash functions, symmetric encryption schemes, public-key systems, etc.) all have security games similar to our biased coin game. Generally, a security game G is played between the challenger and an attacker, A. A corresponds to our distinguisher $D : \{0,1\}^k \to \{0,1\}$, except that A is allowed to be probabilistic. For example, if we allow our biased coin distinguisher to be probabilistic then it is of the form $D : \{0,1\}^k \to [0,1]$ with $D(\omega)$ denoting the probability of guessing the coin is biased given outcome ω. Then, the setup of the security game defines, for every attacker A, a function $f_A : \mathcal{K} \to \mathbb{R}$, where $\mathcal{K}$ is the "key space" (corresponds to our distribution of p). $f_A(k)$ is the expected outcome of the game for a given key K over the coin tosses of A (this corresponds to Pr[D succeeds]). The ideal security of the cryptographic system is then $\epsilon := \mathbb{E}[f_A(U_\mathcal{K})]$, where $U_\mathcal{K}$ is the uniform distribution on the finite set key space $\mathcal{K}$. Notice here we explicitly require the distribution of keys to be uniform. This is the standard assumption for security games, and it is one goal of this paper to explore weaker versions of this expectation, when we replace $U_\mathcal{K}$ with some other distributions on $\mathcal{K}$, which will be called "weak expectation."

For example, take BCG_1. Consider $\mathcal{K} = \{0,1\}$, which is to say, the challenger either takes $p = 0$ or $p = 1$ with equal probability. Then, no matter what distinguisher $D : \{0,1\} \to \{0,1\}$ is used, $\mathbb{E}_p[f_D(p)] = \mathbb{E}_p[\Pr[\text{D succeeds} \mid p]]$ is $\frac{1}{2}$ (we will let the reader check this calculation). In other words, the best we, the attacker, can do is randomly guess!

The notion of distinguishability also offers us tools to tackle other problems. Consider this seemingly unrelated problem: we are given a biased coin, and we want to simulate a fair coin. We again assume that each toss of the coin is independent and identically distributed with probability of heads p. One classic trick is von Neumann's trick [vN51]: we take two tosses and output 1 if the outcome is 10 and 0 if the outcome is 01. We simply ignore the cases where we get 11 or 00. This ensures that we output 0 and 1 with the same probability. However, we will fail with probability $1 - 2(1-p)p$. Let us restrict ourselves to algorithms that do not fail. Can we still make our output distribution uniform? Utilizing our notion of (in)distinguishability: can we make sure that our simulated coin is "indistinguishable" from a fair one with some parameter ϵ? In other words, can we find a function $F : \{0,1\}^n \to \{0,1\}^m$, such that $\Delta(F(C_1^{(n)}), C_0^{(m)}) \leq \epsilon$ for some m and ϵ? The answer is yes. In fact, we will do this in a more general setting than biased coins.

1.2 Overview

From the biased coin example, we glimpsed into the theory of pseudorandomness and (in)distinguishability. In the following sections, we will build the framework for working with pseudorandomness and weak randomness. Roughly speaking, pseudorandomness is

indistinguishable from ideal (uniform) randomness, or more formally, they are close in a semi-metric space of all distributions on some finite set to the uniform distribution, with respect to some class of distinguishers. Weak randomness referes to a distribution with at least some measure of entropy, or denseness relative to the uniform distribution. We will focus on the task of turning weak randomness into pseudorandomness under various assumptions. In Section 2, we formally introduce the tools and build the framework. In Section 3.1, we show the impossibility result using one general weak source. In Section 3.2, we develop the theory of seeded extractors, and in Section 3.3, we focus on two-source extractors.

2 A Model of Randomness

2.1 Notation and Basic Facts

Given a positive integer n, we define $[n] = \{1, ..., n\}$, and use uppercase letters to denote the exponential base 2 of the lowercase letters (for example, $N = 2^n$). We fix a finite universe Ω, and whenever we define a distribution, the universe shall be implicit. Usually, $\Omega = \{0,1\}^n$ (and $|\Omega| = 2^n = N$). A distribution on $\{0,1\}^n$, or the set of n-bit strings, is also called a *source*. Since we are working with a finite sample space, a distribution is uniquely determined given a probability mass function, and vice versa. Hence, we can identify a distribution X on Ω with the probability mass $p_X : \Omega \to [0,1]$. We will also use X and p_X interchangeably, but if X is evaluated as p_X, we will use $X[x] := p_X(x)$ to avoid confusion. Let Y be another distribution (possibly correlated with X), we use (X,Y) to denote their joint distribution. In general, any function $F : \Omega \to [0,1]$ is called a bounded function (on Ω). In particular, a distribution on Ω is a bounded function, and a distinguisher $D : \Omega \to \{0,1\}$ is also a bounded function. The set of all functions from Ω to some set S is denoted S^Ω, e.g. $[0,1]^\Omega$. For a set S, we let U_S denote the uniform distribution on S, and we use U_n to denote $U_{\{0,1\}^n}$, the uniform n-bit strings. Recall that by ideal randomness we mean U_n for some positive integer n. We also use the standard inner product and norms on $\mathbb{R}^\Omega$ (note that $[0,1]^\Omega$ is not a subspace). That is, for $X, Y \in \mathbb{R}^\Omega$:

$$\langle X, Y\rangle = \sum_{\omega\in\Omega} X(\omega)Y(\omega),$$

$$\|X\|_p = (\sum_{\omega\in\Omega} |X(\omega)|^p)^{\frac{1}{p}}, \quad p \geq 1$$

$$\|X\|_\infty = \max_{\omega\in\Omega} |X(\omega)|.$$

Suppose $|\Omega| = N$, recall that we have the following important inequalities on norms,

$$\|X\|_2 \leq \|X\|_1 \leq \sqrt{N}\|X\|_2,$$

$$\|X\|_\infty \leq \|X\|_1 \leq N\|X\|_\infty,$$

$$\|X\|_\infty \leq \|X\|_2 \leq \sqrt{N}\|X\|_\infty,$$

as well as the Cauchy-Schwarz inequality

$$|\langle X, Y\rangle| \leq \|X\|_2\|Y\|_2.$$

Let X be a probability distribution on Ω with probability mass function p_X and $D \in [0,1]^\Omega$. Since we use X and p_X interchangeably, we have:

$$\mathbb{E}[D(X)] = \sum_{\omega\in\Omega} p_X(\omega)D(\omega) = \langle X, D\rangle.$$

Let X' be an i.i.d copy of X. We define the *collision probability*, and *maximum probability* of X as follows,

$$CP(X) = \Pr[X = X'] = \|X\|_2^2,$$
$$MP(X) = \max_{\omega\in\Omega} \Pr[X = \omega] = \|X\|_\infty.$$

Let $F = \{f_1, ..., f_t\}$ be a finite subset of a vector space, for example $\mathbb{R}^\Omega$. The *convex hull* of S is defined to be

$$\mathrm{Conv}(F) = \{\sum_{i=1}^{t} \alpha_i f_i \mid \alpha_i \geq 0, \sum_{i=1}^{t} \alpha_i = 1\}.$$

2.2 Measures of Entropy

The earliest definition of entropy is due to Shannon, and it is very useful in information theory. However, for our applications it is more natural to use the *minimum entropy* and *collision entropy*. Here, we define Rényi entropy, which generalizes the two notions.

Definition 1. *Let X be a distribution, for $\alpha > 1$, the Rényi entropy of X is defined to be*

$$H_\alpha(X) = \frac{\alpha}{1-\alpha}\log\|X\|_\alpha.$$

We also define $H_\infty(X)$ to be the limit of $H_\alpha(X)$ as $\alpha \to \infty$.

In particular, $H_2(X) = -2\log\|X\|_2 = -\log CP(X)$ is called the *collision entropy* of X and $H_\infty(X) = -\log\|X\|_\infty = -\log MP(X)$ is called the *minimum entropy* of X. This definition gives us the exact relation between various measures of entropy and the norm, which would allow us to apply Cauchy-Schwarz inequality.

Proposition 1. *Let X be a distribution. Then $H_\infty(X) \leq H_2(X) \leq 2H_\infty(X)$.*

Proof. First,

$$\begin{aligned} H_2(X) &= -\log\sum_{\omega\in\omega} X[\omega]^2 \\ &\geq -\log\sum_{\omega\in\omega} \|X\|_\infty X[\omega] \\ &= -\log\|X\|_\infty \\ &= H_\infty(X). \end{aligned}$$

Also, since $\|X\|_\infty \leq \|X\|_2$,

$$2H_\infty(X) = -2\log\|X\|_\infty \geq -2\log\|X\|_2 = H_2(X).$$

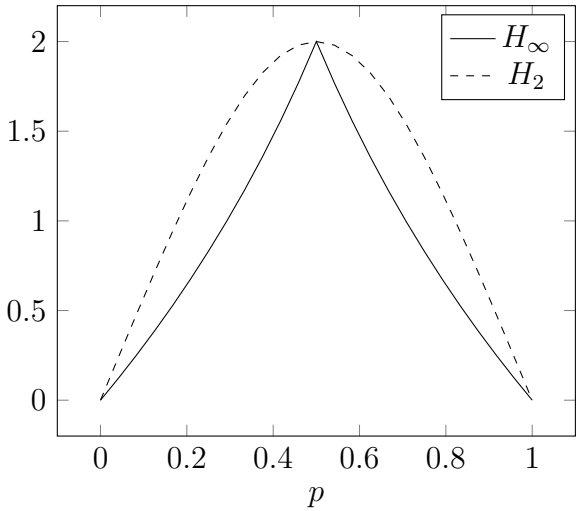
□

Example. Suppose we have a biased coin, C, with $p > \frac{1}{2}$ probability of heads. Let $C^{(k)}$ be k tosses of the coin (viewed as a joint distribution). Again, assume that each toss is independent of the other. Then,

$$\begin{aligned} H_\infty(C^{(k)}) &= -\log\|C^{(k)}\|_\infty \\ &= -\log p^k \\ &= -k\log p. \end{aligned}$$

$$\begin{aligned} H_2(C^{(k)}) &= -\log\|C^{(k)}\|_2^2 \\ &= -\log\sum_{\omega\in\{0,1\}^k} p(\omega)^2 \\ &= -\log\sum_{i=0}^{k}\binom{k}{i}(p^i(1-p)^{k-i})^2. \end{aligned}$$

In particular, if we take $k = 2$, we can plot $H_2(C^{(2)})$ and $H_\infty(C^{(2)})$ as follows:



We can see that $H_\infty(C^{(2)}) \leq H_2(C^{(2)})$ and $H_\infty(C^{(2)}) = H_2(C^{(2)})$ at $p = 0, \frac{1}{2}$, and 1.

2.3 Modelling Weak Sources

Standard cryptography systems assume availability of ideal randomness. Usually, their security is only valid if the randomness is truly uniformly random. However, ideal randomness, as the name suggests is ideal and might not be available to a real system. A more realistic setting is to assume sources with certain entropy bounds on H_2 or H_∞. For a distribution X on $\{0,1\}^n$ with $H_\infty(X) \geq k$, we say that X is a $(n,k)_\infty$-source. Similarly, if $H_2(X) \geq k$, we say that X is a $(n,k)_2$-source. Additionally, when the length of the distribution is clear, we omit n and write k_∞-source or k_2-source.

Definition 2. *Let X be a distribution on $\{0,1\}^n$, we say that X is a flat (n,k)-source for some $0 < k \leq n$, if $\Pr[X = \omega]$ is either 0 or 2^{-k}.*

Proposition 2. *Let X be a $(n,k)_\infty$-source. Then, $X \in \mathrm{Conv}(\mathcal{K}_n)$, where $\mathcal{K}_n$ is the set of all flat (n,k)-sources.*

Proof. We interpret distributions over $\{0,1\}^n$ as vectors over $\mathbb{R}^N$ (recall that $N = 2^n$). We note that the set of (n,k)-sources form a convex polygonal region with exactly the flat (n,k)-sources as its vertices. Hence, $\mathrm{Conv}(\mathcal{K}_n)$ is exactly the set of (n,k)-sources. □

2.4 Semi-Metrics on Distributions

In this section, we formally define the idea of *distinguishability*. First, we fix a class of bounded distinguishers, $\mathcal{D} \subseteq [0,1]^\Omega$. We usually encounter two kinds of distinguishers: *boolean* distringuishers and *bounded* distinguishers. A boolean distinguisher is any function in $\{0,1\}^\Omega$ and bounded distinguisher is any function in $[0,1]^\Omega$. Boolean distinguishers can be thought as the set of deterministic distinguishers. Bounded distinguishers can be thought as the set of probabilistic distinguishers. Hence, to focus on deterministic circuits, we sometimes restrict ourselves to boolean distinguishers. Fixing a subset of boolean functions is also natural in circuit complexity theory. For example, the class of boolean distinguishers could be all boolean functions computable by a circuit of certain size consisting of only **AND**, **OR**, and **NOT** gates with bounded fan-in.

Definition 3. *Let X, Y be distributions and $\mathcal{D} = [0,1]^\Omega$ be a class of distinguishers. The distance (or distinguishability) with respect to $\mathcal{D}$ of X and Y is defined to be*

$$\begin{aligned}\Delta_{\mathcal{D}}(X,Y) &:= \max_{D \in \mathcal{D}} |\langle X - Y, D\rangle| \\ &= \max_{D \in \mathcal{D}} |\mathbb{E}[D(X)] - \mathbb{E}[D(Y)]|.\end{aligned}$$

In addition, if $\mathcal{D}$ is the set of all boolean distinguishers, $\{0,1\}^\Omega$, we simply write $\Delta(X,Y)$, which is called the statistical distance *between X and Y. Further, if $\Delta_{\mathcal{D}}(X,Y) \leq \epsilon$, then we call X and Y ϵ-indistinguishable by $\mathcal{D}$. For notational convenience, we will use the following notation*

$$\Delta_{\mathcal{D}}(X, Y \mid Z) := \Delta_{\mathcal{D}}((X,Z),(Y,Z)).$$

Proposition 3. *Fix a class of distinguishers, $\mathcal{D} \subseteq [0,1]^\Omega$. Then Δ_D is a semi-metric, that is for distributions X, Y, Z over Ω,*

1. $\Delta_{\mathcal{D}}(X, X) = 0$
2. $\Delta_{\mathcal{D}}(X, Y) \geq 0$
3. $\Delta_{\mathcal{D}}(X, Y) = \Delta_{\mathcal{D}}(Y, X)$
4. $\Delta_{\mathcal{D}}(X, Z) \leq \Delta_{\mathcal{D}}(X, Y) + \Delta_{\mathcal{D}}(Y, Z)$

Proof. 1, 2, and 3 easily follow from the definition. We check that $\Delta_{\mathcal{D}}$ satisfy the triangle inequality:

$$\begin{aligned}
\Delta_{\mathcal{D}}(X, Z) &= \max_{D \in \mathcal{D}} |\langle X - Z, D\rangle| \\
&= \max_{D \in \mathcal{D}} |\langle (X - Y) + (Y - Z), D\rangle| \\
&\leq \max_{D \in \mathcal{D}} (|\langle (X - Y), D\rangle| + |\langle (Y - Z), D\rangle|) \\
&\leq \Delta_{\mathcal{D}}(X, Y) + \Delta_{\mathcal{D}}(Y, Z).
\end{aligned}$$

□

Remark. $\Delta_{\mathcal{D}}$ is a semi-metric since it is possible that $\Delta_{\mathcal{D}}(X, Y) = 0$ for $X \neq Y$.

Lemma 1. *Let X, Y be distributions and $\mathcal{D} = [0,1]^\Omega$ be a finite class of bounded distinguishers. Then*

$$\Delta_{\mathrm{Conv}(\mathcal{D})}(X, Y) = \Delta_{\mathcal{D}}(X, Y).$$

Proof. Clearly, $\Delta_{\mathcal{D}}(X, Y) \leq \Delta_{\mathrm{Conv}(\mathcal{D})}(X, Y)$, since $\Delta_{\mathcal{D}} \subseteq \Delta_{\mathrm{Conv}(\mathcal{D})}$.
Now, let $D = \sum_{i=1}^t \alpha_i D_i \in \mathrm{Conv}(\mathcal{D})$, for $\alpha_i \geq 0$ and $\sum_{i=1}^t \alpha_i = 1$. Let $j = \arg\max_j |\langle X - Y, D_j\rangle|$, then

$$\begin{aligned}
|\langle X - Y, D\rangle| &= |\langle X - Y, \sum_{i=1}^t \alpha_i D_i\rangle| \\
&= |\sum_{i=1}^t \alpha_i \langle X - Y, D_i\rangle| \\
&\leq |\langle X - Y, D_j\rangle| \sum_{i=1}^t \alpha_i \\
&= |\langle X - Y, D_j\rangle|.
\end{aligned}$$

□

Notice that $\{0,1\}^\Omega$ is finite, hence we obtain the following corollary.

Corollary. *Let X, Y be distributions. Then*

$$\Delta_{[0,1]^\Omega}(X,Y) = \Delta_{\{0,1\}^\Omega}(X,Y).$$

If we take $\mathcal{D}$ to be the set of all boolean distinguishers, then the distance is minimized. We note that in this case, the statistical distance is exactly half the distance induced by the 1-norm, as given in the following lemma.

Lemma 2. *Let X, Y be distributions on Ω. Then,*

$$\Delta(X,Y) = \frac{1}{2}\|X - Y\|_1.$$

Proof. First, let $T = \{t \in \Omega \mid p_X(t) > p_Y(t)\}$ and $D : \Omega \to \{0,1\}$ be such that $D(\omega) = 1 \iff \omega \in T$. We compute that

$$\begin{aligned}
\|X - Y\|_1 &= \sum_{\omega\in\Omega} |X[\omega] - Y[\omega]| \\
&= \sum_{\omega\in T} X[\omega] - Y[\omega] + \sum_{\omega\in\overline{T}} Y[\omega] - X[\omega] \\
&= \left|\sum_{\omega\in T} X[\omega] - Y[\omega]\right| + \left|(1 - \sum_{\omega\in T} Y[\omega]) - (1 - \sum_{\omega\in T} X[\omega])\right| \\
&= 2|\mathbb{E}[D(X) = 1] - \mathbb{E}[D(Y) = 1]| \\
&= 2\Delta_D(X,Y).
\end{aligned}$$

Let $D' \in \{0,1\}^\Omega$. Consider $(2D' - 1)$. Check that $\|2D' - 1\|_\infty = 1$ and $\langle X - Y, 2D' - 1\rangle = 2\langle X - Y, D'\rangle$. Hence, we have that

$$\begin{aligned}
2|\langle X - Y, D'\rangle| &= |\langle X - Y, 2D' - 1\rangle| \\
&= \left|\sum_{\omega\in\Omega} (X - Y)[\omega](2D - 1)[\omega]\right| \\
&\leq \|2D - 1\|_\infty \sum_{\omega\in\Omega} |(X - Y)[\omega]| \\
&= \|X - Y\|_1.
\end{aligned}$$

□

Lemma 3. *Let (X, X') be jointly distributed and let (Y, Y') be jointly distributed over the same sample space, $\Omega \times \Omega'$. Then,*

$$\Delta((X,X'),(Y,Y')) \geq \Delta(X,Y).$$

Proof. Take any $D : \Omega \to \{0,1\}$, and let

$$D' : \Omega \times \Omega' \to \{0,1\};$$

$$D'(x,x') = D(x).$$

We compute that,

$$\begin{aligned}\Delta_{D'}((X,X'),(Y,Y')) &= |\Pr[D'(X,X')=1] - \Pr[D'(Y,Y')=1]| \\ &= |\Pr[D(X)=1] - \Pr[D(Y)=1]| \\ &= \Delta_D(X,Y).\end{aligned}$$

□

3 Randomness Extractors

In this section, we will allow all possible distinguishers and use the notion of statistical distance, which was shown to be equivalent to the 1-norm. In particular, we show how to convert weak sources to statistically close to uniform.

Recall that in Section 1.1 we raised the question of converting biased coin flips to fair coin flips. Let us fix some $\alpha \in \{2,\infty\}$. We will relax our input to be some $(n,k)_\alpha$-source, X. Can we extract almost uniform bits from X? In other words, for $n > m$, can we find a function, $F : \{0,1\}^n \to \{0,1\}^m$, such that for any $(n,k)_\alpha$-source X, $\Delta(F(X),U_m) < \epsilon(n,k,m)$? As it turns out, it cannot be done with a deterministic function; but if we allow ourselves to pick uniformly from a collection of functions, we can achieve such task. Furthermore, if we are given two independent weak sources with sufficient amount of entropy, then deterministic extraction becomes possible again.

3.1 Impossibility Result for Single Source

Suppose we have a function $f : \{0,1\}^n \to \{0,1\}$. For notational convenience, for a boolean function f, we will use $M(f)$ to denote the larger of the two sets: $f^{-1}(0), f^{-1}(1)$ (if they are the same size either will do). Hence, we have that $|M(f)| \geq 2^{n-1}$. Consider $X = U_{M(f)}$. By construction, X is a $(n, n-1)_\infty$ source and $f(X)$ is constant. In fact, this can be generalized to $(n,k)_\infty$-block sources—we cannot even extract one bit from them.

Proposition 4. *Let $f : \{0,1\}^n \to \{0,1\}^m$ be any function, then there exists an $(n, n-1)_\infty$ source X such that $\Delta(f(X),U_m) = \frac{1}{2}$.*

Proof. We let f_1 be the function that restricted to only the first bit of f. By the argument above, $\Delta(f_1(X),U_1) = \frac{1}{2}$. Hence, by Lemma 3, $\Delta(f,U_m) = \frac{1}{2}$. □

The upshot is that if we are only given one weak source, we cannot extract good randomness from it deterministically. This is because the source can be generated adversarially based on our extractor.

3.2 Seeded Extractors

If we are allowed some additional ideal random seed, we are then able to extract good randomness from a weak source. Keep in mind that the reason we want to extract from a weak source is that perfect randomness might not be available, hence this really does not help us much in the setting where we cannot obtain ideal randomness. The nature of the game will be to try to keep the random seed needed *small*, this way the dependence of ideal randomness is minimized. For some applications, we want the seed to be $O(\log n)$ to extract from a (n, k) source, and such constructions have been done. However, we will not mention them here. Instead, we will prove what is called the Leftover Hash Lemma and construct an iterative scheme to extract from long sources.

Definition 4. *[NZ96] Let $\alpha \in \{2, \infty\}$. A function of the form $Ext : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $(k, \epsilon)_\alpha$-randomness extractor (Ext) if, for K uniformly random in $\{0,1\}^d$ and any $(n, k)_\alpha$-source X,*

$$\Delta(Ext(X, K), U_m) \leq \epsilon.$$

Furthermore, we say that the extractor is strong if

$$\Delta(Ext(X, K), U_m \mid K) \leq \epsilon.$$

We will construct a well-known family of $(k, \epsilon)_2$ extractors using universal independent (hash) functions from [CW79]. Notice that since $H_2(X) \geq H_\infty(X)$, any $(k, \epsilon)_2$ extractor is a $(k, \epsilon)_\infty$ extractor.

3.2.1 Universal Hash Functions

Definition 5. *A collection of functions*

$$\boldsymbol{H} = \{H_s : \{0,1\}^n \to \{0,1\}^m \mid s \in \{0,1\}^d\}$$

is universal, if for H uniformly chosen in $\boldsymbol{H}$, and two distinct input $x \neq y$,

$$\Pr(H(x) = H(y)) = \frac{1}{M} = \frac{1}{2^m}.$$

For notational convenience, we also call H, the uniform sample, a universal hash function.

One elegant construction comes from the inner product structure on $\mathbb{F}_{2^n}$.

Construction 1. For any $m, l \in \mathbb{N}$, let $n = ml$. We identify $\{0,1\}^m$ as $\mathbb{F}_{2^m}$ (binary field of size 2^m) and $\{0,1\}^n$ as $\mathbb{F}_{2^m}^l$ (l-dimensional vector space over binary field of size 2^m). We consider the inner product function (standard dot-product of vectors), $\mathrm{IP} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$,

$$\mathrm{IP}(x, y) := x \cdot y.$$

Proposition 5. *For any $m, l \in \mathbb{N}$,*

$$\boldsymbol{H} = \{H_v = \mathrm{IP}(-, v) \mid v \in \mathbb{F}_{2^m}^l\}$$

is an universal collection of hash functions.

Proof. Let v be uniform in $\mathbb{F}_{2^m}^l$. Let $x = (x_1, ..., x_l) \neq y = (y_1, ..., y_l) \in F$. Suppose $x_j \neq y_j$. Then,

$$\begin{aligned}
&\Pr[H_v(x) = H_v(y)] \\
&= \Pr[\sum_{i=1}^{l} v_i x_i = \sum_{i=1}^{l} v_i y_i] \\
&= \Pr[v_i(x_j - y_j) = \sum_{\substack{i=1,...,l \\ i \neq j}} v_i(x_i - y_i)] \\
&= \frac{1}{M}.
\end{aligned}$$

□

Lemma 4. *Let X be a distribution over $\{0,1\}^n$. Then,*

$$\|X - U_d\|_2^2 = \|X\|_2^2 - \frac{1}{N}.$$

Proof. Compute that

$$\begin{aligned}
\|X - U_n\|_2^2 &= \|X\|_2 + \|U_n\|_2 - 2\langle X, U_n\rangle \\
&= \|X\|_2 - \frac{1}{N}.
\end{aligned}$$

□

Lemma 5. *Let $f : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a function such that $f(-, U_d)$ is universal, then*

$$\|(U_d, f(X, U_d)) - (U_d, U_m)\|_2^2 \leq \frac{CP(X)}{D}.$$

Proof. First, we bound the collision probability of $(U_d, f(X, U_d))$. Let s_1, s_2 be independently uniform over $\{0,1\}^d$ and x_1, x_2 be two independent copies of X. Observe that

$$\begin{aligned}
&CP((U_d, f(X, U_d))) \\
&= \Pr[(s_1, f(x_1, s_1)) = (s_2, f(x_2, s_2))] \\
&= \Pr[s_1 = s_2] \cdot \Pr[f(x_1, s_1) = f(x_2, s_2) \mid s_1 = s_2] \\
&= \frac{1}{D}(\Pr[x_1 = x_2] + \Pr[x_1 \neq x_2] \Pr[f(x_1, s_1) = f(x_2, s_1) \mid x_1 \neq x_2]) \\
&\leq \frac{1}{D}(CP(X) + \frac{1}{M}).
\end{aligned}$$

Now, by Lemma 4,

$$\begin{aligned} &\|(U_d, f(X, U_d)) - (U_d, U_m)\|_2^2 \\ &= CP((U_d, f(X, U_d))) - \frac{1}{DM} \\ &= \frac{CP(X)}{D}. \end{aligned}$$

□

Theorem 1 (**Leftover Hash Lemma** [ILL89])**.** *Let X be a $(n, k)_2$ source. Let $\boldsymbol{H} = \{H_s : \{0,1\}^n \to \{0,1\}^m \mid s \in \{0,1\}^d\}$ be an universal collection of functions. Then, the extractor defined by*

$$Ext(x, s) = h_s(x),$$

is a strong $(k, \epsilon)_2$-randomness extractor with $\epsilon = \frac{1}{2}\sqrt{\frac{M}{K}} = \frac{1}{2}\sqrt{2^{n-k}}$.

Proof. Let $v = (U_d, Ext(X, U_d)) - (U_d, U_m)$. We can bound the statistical distance by using Lemma 5.

$$\begin{aligned} \Delta((U_d, Ext(X, U_d)), (U_d, U_m)) &= \frac{1}{2}\|v\|_1 \\ &\leq \frac{\sqrt{DM}}{2}\|v\|_2 \\ &\leq \frac{\sqrt{DM}}{2}\sqrt{\frac{CP(X)}{D}} \\ &= \frac{1}{2}\sqrt{\frac{M}{K}}. \end{aligned}$$

□

Hence, for any $m, l \geq 1$, let $n = ml$. Then, IP, from Construction 1, is universal with seed length n, input length n, and output length m. Hence, we have that $Ext(x, s) = h_s(x)$ is a $(k, \frac{1}{2}\sqrt{\frac{M}{K}})_2$-extractor. However, we need n uniform random bits for this task! For example, suppose we want 256 uniform random bits of error at most 2^{-128} ($m = 256$ and $\epsilon = 2^{-128}$). As a result, we need $k \geq 512$, and n need to be more than k for the source to be actually weak. Hence, we would at least a 768-bit uniform seed. The main caveat in this construction is that we need as long an initial seed as the length of the weak source. What if the weak source we want to extract from is very long? For instance, we could be measuring the CPU temperature and we are guaranteed to get some entropy each time we measure it. In such cases, the resulting weak source can be very long but will have good accumulative entropy.

3.2.2 Almost Universal Hash Functions

In this section, we consider reducing the seed length relative to the weak source length to the extractor, or equivalently the universal hash function construction. We will do this by relaxing the requirement of being *exactly* universal to *almost* universal. Then, we will prove a slightly modified version of Leftover Hash Lemma for almost universal hash functions.

Definition 6. *A collections of functions*

$$\boldsymbol{H} = \{H_s : \{0,1\}^n \to \{0,1\}^m \mid s \in \{0,1\}^d\}$$

is δ-almost universal, if for any $x \neq y \in \{0,1\}^n$,

$$\Pr[H(x) = H(y)] \leq \frac{1+\delta}{M}.$$

If $\delta = 0$, we will say that $\boldsymbol{H}$ is a universal function family.

It turns that we if we are given a source with high enough minimum entropy, we can get away with almost universal functions in place of pairwise universal functions. The length of the initial seed required, however, can be much shorter than the input length for almost universal functions. We prove an extension of Lemma 5 for almost universal hash functions.

Lemma 6. *Let $f : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a function such that $f(-, U_d)$ is δ-almost universal. Then*

$$\|(U_d, f(X, U_d)) - (U_d, U_m)\|_2^2 \leq \frac{CP(X)}{D} + \frac{\delta}{DM}$$

Proof. First, we bound the collision probability of $(U_d, f(X, U_d))$, let s_1, s_2 be independently uniform over $\{0,1\}^d$ and x_1, x_2 be two independent copies of X. Then

$$\begin{aligned}
&CP((U_d, f(X, U_d))) \\
&= \Pr[(s_1, f(x_1, s_1)) = (s_2, f(x_2, s_2))] \\
&= \Pr[s_1 = s_2] \cdot \Pr[f(x_1, s_1) = f(x_2, s_2) \mid s_1 = s_2] \\
&= \frac{1}{D}(\Pr[x_1 = x_2] + \Pr[x_1 \neq x_2] \Pr[f(x_1, s_1) = f(x_2, s_1) \mid x_1 \neq x_2]) \\
&\leq \frac{1}{D}(CP(X) + \frac{1+\delta}{M}).
\end{aligned}$$

Now, by Lemma 4,

$$\begin{aligned}
&\|(U_d, f(X, U_d)) - (U_d, U_m)\|_2^2 \\
&= CP((U_d, f(X, U_d))) - \frac{1}{DM} \\
&= \frac{CP(X)}{D} + \frac{\delta}{DM}.
\end{aligned}$$

□

A modified version of the Leftover Hash Lemma follows naturally.

Theorem 2 (Leftover Hash Lemma for Almost Universal Functions)**.** *Let X be a $(n,k)_2$ source. Let $\boldsymbol{H} = \{h_s : \{0,1\}^n \to \{0,1\}^m \mid s \in \{0,1\}^d\}$ be a δ-almost universal collection of functions. Then, the extractor defined by*

$$Ext(x,s) = h_s(x),$$

is a strong $(k,\epsilon)_2$-randomness extractor with $\epsilon = \frac{1}{2}\sqrt{\frac{M}{K} + \delta}$.

Proof. Let $v = (U_d, Ext(X,U_d)) - (U_d, U_m)$, then by Lemma 6,

$$\begin{aligned}
\Delta((U_d, Ext(X,U_d)),(U_d,U_m)) &= \frac{1}{2}\|v\|_1 \\
&\leq \frac{\sqrt{DM}}{2}\|v\|_2 \\
&\leq \frac{\sqrt{DM}}{2}\sqrt{\frac{CP(X)}{D} + \frac{\delta}{DM}} \\
&= \frac{1}{2}\sqrt{\frac{M}{K} + \delta}.
\end{aligned}$$

□

Next, we show that polynomial maps are almost universal functions. In addition, they are computable by iteration! This allow us to slowly accumulate entropy into a system. Actually, the entropy input can be arbitrarily small at each step. But we insist that the sequence of entropy input should eventually exceed a threshold k at most after some t steps. Of course we will assume that we have access to some initial randomness.

Construction 2 ([DPR+13])**.** For any positive integers m, d, t, such that $0 < m \leq d$, we construct a family of functions with input length $n = td$, seed length $2d$, and output length m. We identify $\{0,1\}^d$ as $\mathbb{F}_{2^d}$. We will break td length input into t chunks, $I_1, \ldots, I_t$, each of length d, and compute the output iteratively as follows. We have seed $(X, X') \in \{0,1\}^{2d}$. For any $0 \leq i \leq t$, the state S at time i is $S_i \in \{0,1\}^d$. X, X' are initialized uniformly randomly and S_0 is initialized to $1 \in \mathbb{F}_{2^d}$. At step $i \geq 1$, we take some entropy I_i and set $S_i = S_{i-1} \cdot X + I_i$, where the operations are done in $\mathbb{F}_{2^d}$. It is easy to see that at step i,

$$S_t = X^i + I_1 X^{i-1} + I_2 X^{i-2} + \cdots + I_i,$$

which is a polynomial in X of degree i. The output, at step t, is the first m-bit of $S_t X'$, which we denote as $[S_t X']_1^m$. In other words, at step t, our hash function outputs

$$H_{X,X'}(I_1, \ldots, I_t) = [X'(X^t + I_1 X^{t-1} + I_2 X^{t-2} + \cdots + I_t)]_1^m.$$

Proposition 6.

$$H_{X,X'}(I_1,\ldots,I_t) = [X'(X^t + I_1X^{t-1} + I_2X^{t-2} + \cdots + I_t)]_1^m$$

is $t2^{m-d}$-almost universal with seed length $2d$, input length $n = td$, and output length m.

Proof. Let $(a_1,\ldots,a_t) \neq (b_1,\ldots,b_t)$ be two different inputs. Compute that

$$\begin{aligned}
&\Pr[H_{X,X'}(a_1,\ldots,a_t) = H_{X,X'}(b_1,\ldots,b_t)] \\
&= \Pr[[(a_1X^{t-1} + a_2X^{t-2} + \cdots + a_t)X']_1^m = [(b_1X^t + b_2X^{t-2} + \cdots + b_t)X']_1^m] \\
&= \Pr[\sum_{i=1}^{t} a_iX^{t-i} = \sum_{i=1}^{t} b_iX^{t-i}] + \Pr[[AX']_1^m = [BX']_1^m] \\
&= \Pr[\sum_{i=1}^{t} (a_i - b_i)X^{t-i} = 0] + \frac{1}{M} \\
&\leq \frac{t}{2^d} + \frac{1}{M} \\
&= \frac{1 + t2^{m-d}}{M}
\end{aligned}$$

The last inequality follows from the fact that $\sum_{i=1}^{t}(a_i - b_i)X^{t-i}$ is a polynomial of degree at most t in X, therefore it has most t roots. □

We plug $\delta = t2^{m-d}$ into $\epsilon = \frac{1}{2}\sqrt{\frac{M}{K} + \delta}$, we get

$$4\epsilon^2 = \frac{2^m}{2^k} + t\frac{2^m}{2^d}.$$

Solving for k, we get that

$$k = m - \log(4\epsilon^2 - t2^{m-d}).$$

As an example, take $m = 256$, $d = 512$ and set desired $\epsilon = 2^{-128}$. Then, we would need accumulative minimum entropy of $I_1, ..., I_t$ to be at least $256 - \log(4 \times 2^{-256} + t2^{-256}) = 512 - \log(4+t)$. We see that the growth of entropy needed is only logarithmic in t. However, we still need 1024 bits of initial randomness to achieve extraction of 256 bits.

3.3 Two-Source Extractors

Can we relax the requirement for an ideal random seed? We still want our seed to be independent of the weak source, otherwise the impossibility results from Section 3.1 would apply. One natural relaxation is then to assume the availability of *two* weak sources that are *independent*.

Definition 7. *Let $\alpha \in \{2, \infty\}$. A function of the form $Ext : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $(k, k', \epsilon)_\alpha$-two source randomness extractor if, any $(n,k)_\alpha$-source X and any $(d,k')_\alpha$-source Y,*

$$\Delta(Ext(X,Y), U_m) \leq \epsilon.$$

Furthermore, the extractor is strong (in Y) if

$$\Delta(Ext(X,Y), U_m \mid Y) \leq \epsilon.$$

We abbreviate $(k,k,\epsilon)_\alpha$ as simply $(k,\epsilon)_\alpha$.

Remark. A strong (k,n,ϵ)-two-source extractor, with seed length n, is also a strong seeded extractor. We can further assume the availability of k independent weak sources for $k > 2$. Such extractors are called *independent-source extractors* but we will not consider them here.

3.3.1 Seeded Extractors with Weak Seeds

In fact, any strong seeded extractor, for instance Construction 1 and 2, is a two-source extractor with weaker error. Here, we present a new argument matching the bound in [Rao07], which, at the same time, applies to all collision entropy extractors. We were unable to find such results obtained via similar techniques in the literature. In [CG88], the same results were obtained only for the inner product function with one output bit and proven using Lindsey's Lemma. We begin with the following lemma.

Lemma 7. *Let $f : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be any function, and let Y be a $(d,k)_2$ source and X be any distribution on $\{0,1\}^n$. Then,*

$$\|(Y, f(X,Y)) - (Y, U_m)\|_1 \leq \sqrt{\frac{D^2 M}{K}} \|(U_d, f(X,U_d)) - (U_d, U_m)\|_2.$$

Proof.

$$\begin{aligned}
&\|(Y, f(X,Y)) - (Y, U_m)\|_1 \\
&= \sum_{(r,s)\in\{0,1\}^{d+m}} |(Y, f(X,Y))[r,s] - (Y, U_m)[r,s]| \\
&= \sum_{r,s} Y[r] |f(X,r)[s] - U_m[s]| \\
&\leq \sqrt{\sum_{r,s} Y[r]^2} \sqrt{\sum_{r,s} |f(X,r)[s] - U_m[s]|^2} && \text{(Cauchy-Schwarz)} \\
&= \sqrt{D^2 M \sum_r Y[r]^2} \sqrt{\sum_{r,s} |\frac{1}{2^d} f(X,r)[s] - \frac{1}{2^d} U_m[s]|^2} \\
&= \sqrt{\frac{D^2 M}{K}} \|(U_d, f(X,U_d)) - (U_d, U_m)\|_2.
\end{aligned}$$

□

Theorem 3 (**Leftover Hash Lemma for Weak Seeds**). *Suppose that* $f : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is such that* $f(-, U_d)$ *is* δ*-universal, then* f *is a strong (in* Y*)* (k, k', ϵ)*-two-source extractor, with* $\epsilon = \frac{1}{2}\sqrt{\frac{D}{K'}(\frac{M}{K} + \delta)}$.

Proof. We compute that

$$\begin{aligned}
&\Delta((Y, f(X,Y)), (Y, U_m)) \\
&= \frac{1}{2}\|(Y, f(X,Y)) - (Y, U_m)\|_1 \\
&\leq \frac{1}{2}\sqrt{\frac{D^2M}{K'}}\|(f(X, U_d), U_d) - (U_m, U_d)\|_2 && \text{(Lemma 7)} \\
&\leq \frac{1}{2}\sqrt{\frac{D^2M}{K'}}\sqrt{\frac{CP(X)}{D} + \frac{\delta}{DM}} && \text{(Lemma 6)} \\
&\leq \frac{1}{2}\sqrt{\frac{D}{K'}(\frac{M}{K} + \delta)}.
\end{aligned}$$

□

Remark. Notice that when $k' = d$ ($Y = U_d$), the bound matches Theorem 2. If $\delta = 0$ ($f(-, U_d)$ is perfectly universal), then $\epsilon = \frac{1}{2}\sqrt{\frac{DM}{KK'}} = \frac{1}{2} \cdot 2^{\frac{d+m-k-k'}{2}}$, which matches the bound obtained by using character bounds in [Rao07].

We look at an example. Suppose we want a 256-bit secret with error at most $\epsilon = 2^{-128}$. Then it suffices to have $n + 512 = k_1 + k_2$ for two $(n, k_1)_2$ and $(n, k_2)_2$ sources. For example, we can use two $(1024, 768)_2$ sources X, Y.

For Construction 2, plugging $\delta = t\frac{M}{D}$ into $\epsilon = \frac{1}{2}\sqrt{\frac{D}{K'}(\frac{M}{K} + \delta)}$, we get

$$\epsilon = \frac{1}{2}\sqrt{\frac{DM}{KK'} + \frac{tM}{K'}}.$$

Let $m = 256$ and $\epsilon = 2^{-128}$. It suffices to have

$$k + k' - d - m = k' - \log(t) - m = 256.$$

If we assume that $t \leq 2^{128}$, then $k' \geq 640$ suffices. Let us further assume that $d = 1024$. Then, $k \geq 896$ suffice. Hence, we have a two-source extractor for $(1024t, 896)_2$ and $(1024, 640)_2$ source for t upto 2^{128} that extracts 256-bit of ideal randomness with error $\epsilon = 2^{-128}$.

4 Summary and Further Reading

In summary, we introduced the framework for working with weak randomness and pseudo-randomness. We showed two explicit constructions of randomness extractors by constructing

two universal hash functions and proving two versions of the Leftover Hash Lemma. We also presented a new argument to extend the Leftover Hash Lemma to work with weak seeds, which states that universal hash functions are two-source extractors. We point out that the technique might be applicable to other cryptographic objects and systems.

Following initiating work on weak randomness [SV86, CG88], a line of research [NZ96, Tre01, LRVW03, GUV09] set out to reduce the seed length required for seeded extraction, which was first considered in [NZ96]. Optimal logarithmic seed length has been realized [GUV09]. On the other hand, significant research has been put into independent-source extractors [DO03, DEOR04, BKS+05, BIW06, Rao07, Li13, Li15b, CZ15, Li15a] with the goal of constructing two-source extractors for logarithmic entropy. The first two-source extractors for poly-logarithmic entropy was constructed recently [CZ15, Li15a].

A new type of extractors, called *non-malleable* extractors, have also gained attention due to its application to privacy amplification with an active adversary [DW09, Li12, DLWZ14, CRS14]. It has been shown that non-malleable extractors can be constructed from two-source extractors, and vice versa [Li12].

Interestingly, an analog of pseudorandomness and indistinguishability were used in the proof of the Green-Tao theorem [GT04, TZ06], which states that the primes contain arbitrary length arithmetic progressions. One key component of argument was the Dense Model Theorem, which can be stated in terms of our framework [RTTV08].

References

[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/`.

[BIW06] Boaz Barak, Russell Impagliazzo, and Avi Wigderson. Extracting randomness using few independent sources. *SIAM Journal on Computing*, 36(4):1095–1118, 2006.

[BKS+05] Boaz Barak, Guy Kindler, Ronen Shaltiel, Benny Sudakov, and Avi Wigderson. Simulating independence: New constructions of condensers, ramsey graphs, dispersers, and extractors. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 1–10. ACM, 2005.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17(2):230–261, April 1988.

[CRS14] Gil Cohen, Ran Raz, and Gil Segev. Nonmalleable extractors with short seeds and applications to privacy amplification. *SIAM Journal on Computing*, 43(2):450–476, 2014.

[CW79] J.Lawrence Carter and Mark N. Wegman. Universal classes of hash functions. *Journal of Computer and System Sciences*, 18(2):143 – 154, 1979.

[CZ15] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 22, page 119, 2015.

[DEOR04] Yevgeniy Dodis, Ariel Elbaz, Roberto Oliveira, and Ran Raz. Improved randomness extraction from two independent sources. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, pages 334–344. Springer, 2004.

[DLWZ14] Yevgeniy Dodis, Xin Li, Trevor D Wooley, and David Zuckerman. Privacy amplification and nonmalleable extractors via character sums. *SIAM Journal on Computing*, 43(2):800–830, 2014.

[DO03] Yevgeniy Dodis and Roberto Oliveira. On extracting private randomness over a public channel. In *Approximation, Randomization, and Combinatorial Optimization.. Algorithms and Techniques*, pages 252–263. Springer, 2003.

[DPR+13] Yevgeniy Dodis, David Pointcheval, Sylvain Ruhault, Damien Vergnaud, and Daniel Wichs. Security analysis of pseudo-random number generators with input: /dev/random is not robust. Cryptology ePrint Archive, Report 2013/338, 2013.

[DSSDW14] Yevgeniy Dodis, Adi Shamir, Noah Stephens-Davidowitz, and Daniel Wichs. How to eat your entropy and have it too – optimal recovery strategies for compromised rngs. Cryptology ePrint Archive, Report 2014/167, 2014.

[DW09] Yevgeniy Dodis and Daniel Wichs. Non-malleable extractors and symmetric key cryptography from weak secrets. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 601–610. ACM, 2009.

[DY13] Yevgeniy Dodis and Yu Yu. Overcoming weak expectations. In Amit Sahai, editor, *Theory of Cryptography*, volume 7785 of *Lecture Notes in Computer Science*, pages 1–22. Springer Berlin Heidelberg, 2013.

[GT04] Ben Green and Terrance Tao. The primes contain arbitrarily long arithmetic progressions. *ArXiv Mathematics e-prints*, April 2004.

[GUV09] Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. Unbalanced expanders and randomness extractors from parvaresh–vardy codes. *Journal of the ACM (JACM)*, 56(4):20, 2009.

[ILL89] Russell Impagliazzo, Leonid A. Levin, and Michael Luby. Pseudo-random generation from one-way functions. In *Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing*, STOC '89, pages 12–24, New York, NY, USA, 1989. ACM.

[Li12] Xin Li. Non-malleable extractors, two-source extractors and privacy amplification. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 688–697. IEEE, 2012.

[Li13] Xin Li. Extractors for a constant number of independent sources with polylogarithmic min-entropy. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 100–109. IEEE, 2013.

[Li15a] Xin Li. Improved Constructions of Two-Source Extractors. *ArXiv e-prints*, August 2015.

[Li15b] Xin Li. Three-source extractors for polylogarithmic min-entropy. *arXiv preprint arXiv:1503.02286*, 2015.

[LRVW03] Chi-Jen Lu, Omer Reingold, Salil Vadhan, and Avi Wigderson. Extractors: Optimal up to constant factors. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 602–611. ACM, 2003.

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

[Rao07] Anup Rao. An exposition of bourgains 2-source extractor. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 14, 2007.

[RTTV08] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. New Proofs of the Green-Tao-Ziegler Dense Model Theorem: An Exposition. *ArXiv e-prints*, June 2008.

[SV86] Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Sciences*, 33(1):75 – 87, 1986.

[Tre01] Luca Trevisan. Extractors and pseudorandom generators. *Journal of the ACM*, 48(4):860–879, 2001.

[TZ06] Terrance Tao and Tamar Ziegler. The primes contain arbitrarily long polynomial progressions. *ArXiv Mathematics e-prints*, October 2006.

[Vad11] Salil P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(13):1–336, 2011.

[Vaz86] Umesh Virkumar Vazirani. *Randomness, Adversaries and Computation (Random Polynomial Time)*. PhD thesis, University of California, Berkeley, 1986. AAI8718194.

[vN51] John von Neumann. Various techniques used in connection with random digits. *National Bureau of Standards, Applied Math Series*, 12:36–38, 1951.