Computability 1 (2012) 1–29
DOI
IOS Press

Dimension, Pseudorandomness and Extraction of Pseudorandomness*

Manindra Agrawal
Department of Computer Science & Engineering, Indian Institute of Technology Kanpur, Kanpur-208016, India
manindra@cse.iitk.ac.in
http://www.cse.iitk.ac.in/users/manindra/

Diptarka Chakraborty
Department of Computer Science & Engineering, Indian Institute of Technology Kanpur, Kanpur-208016, India
diptarka@cse.iitk.ac.in
http://www.cse.iitk.ac.in/users/diptarka/

Debarati Das
Charles University in Prague, Computer Science Institute of Charles University, Malostranské námesti 25, 118 00 Praha 1, Czech Republic
debaratix710@gmail.com

Satyadev Nandakumar
Department of Computer Science & Engineering, Indian Institute of Technology Kanpur, Kanpur-208016, India
satyadev@cse.iitk.ac.in
http://www.cse.iitk.ac.in/users/satyadev/

Abstract. In this paper we propose a quantification of ensemble of distributions on a set of strings, in terms of how close to pseudorandom a distribution is. The quantification is an adaptation of the theory of dimension of sets of infinite sequences introduced by Lutz. Adapting Hitchcock's work, we also show that the logarithmic loss incurred by a predictor on an ensemble of distributions is quantitatively equivalent to the notion of *dimension* we define. Roughly, this captures the equivalence between pseudorandomness defined via indistinguishability and via unpredictability. Later we show some natural properties of our notion of dimension. We also do a comparative study among our proposed notion of dimension and two well known notions of computational analogue of entropy, namely HILL-type pseudo min-entropy and next-bit pseudo Shannon entropy.

Further, we apply our quantification to the following problem. If we know that the dimension of an ensemble of distributions on the set of n-length strings is $s \in (0,1]$, can we extract out $O(sn)$ pseudorandom bits out of the distribution? We show that to construct such extractor, one needs at least $\Omega(\log n)$ bits of pure randomness. However, it is still open to do the same using $O(\log n)$ random bits. We show that deterministic extraction is possible in a special case - analogous to the bit-fixing sources introduced by Chor *et al.*, which we term *nonpseudorandom bit-fixing source*. We adapt the techniques of Gabizon, Raz and Shaltiel to construct a deterministic *pseudorandom extractor* for this source.

By the end, we make a little progress towards P vs. BPP problem by showing that existence of optimal stretching function that stretches $O(\log n)$ input bits to produce n output bits such that output distribution has dimension $s \in (0,1]$, implies P=BPP.

Keywords: Pseudorandomness, Dimension, Unpredictability, Pseudoentropy, Pseudorandom Extractor

*A preliminary version of this paper [1] was accepted in the 35th IARCS Annual Conference on Foundation of Software Technology and Theoretical Computer Science, FSTTCS 2015, December 16-18, 2015, Bangalore, India

1. Introduction

Incorporating randomness in a feasible computation is one of the basic primitives in theoretical computer science. Fortunately, any efficient (polynomial time) randomized algorithm does not require pure random bits. What it actually needs is a source that *looks* random to it and this is where the notion of *pseudorandomness* [5, 36] comes into picture. Since its introduction, pseudorandomness has been fundamental to the domain of cryptography, complexity theory and computational learning theory. Pseudorandomness is mainly a computational approach to study the nature of randomness, and *computational indistinguishability* [11] played a pivotal role in this. Informally, a distribution is said to be pseudorandom if no efficient algorithm can distinguish it from the uniform distribution. Another way of looking at computational indistinguishability is via the notion of *unpredictability* of distributions, due to Yao [36]. Informally, a distribution is *unpredictable* if there is no efficient algorithm that, given a prefix of a string coming from that distribution, can guess the next bit with a significant success probability. This line of research naturally posed the question of constructing algorithms that can generate pseudorandom distributions, known as *pseudorandom generators*. Till now we know such constructions by assuming the existence of *one-way functions*. It is well known that constructibility of an *optimal pseudorandom generator* implies complete derandomization (i.e., P=BPP) and *exponential hardness assumption* on one-way function enables us to do that. However, Nisan and Wigderson [27] showed that the existence of an exponential *hard function*, which is a much weaker assumption, is also sufficient for this purpose. The assumption was further weakened in [19].

In order to characterize the class of random sources, information theoretic notion of *min-entropy* is normally used. A computational analogue of entropy was introduced by Yao [36] and was based on compression. Håstad, Impagliazzo, Levin and Luby [13] extended the definition of min-entropy in computational settings while giving the construction of a pseudorandom generator from any one-way function. This HILL-type *pseudoentropy* basically extends the definition of pseudorandomness syntactically. Relations among above two types of pseudoentropy was further studied in [4]. A more relaxed notion of pseudoentropy, known as *next-bit Shannon pseudoentropy*, was later introduced by Haitner, Reingold and Vadhan [12] in the context of an efficient construction of a pseudorandom generator from any one-way function. In a follow up work [34], the same notion was alternatively characterized by *KL-hardness*. So far it is not clear which of the above notions is the most appropriate or whether they are at all suitable to characterize distributions in terms of the degree of pseudorandomness in it.

In this paper, we first propose an alternative measure to quantify the amount of pseudorandomness present in a distribution. This measure is motivated by the ideas of *dimension* [24] and *logarithmic loss unpredictability* [16]. Lutz used the betting functions known as *gales* to characterize the *Hausdroff dimension* of sets of infinite sequences over a finite alphabet. The definition given by Lutz cannot be carried over directly, because here we consider the distributions over finite length strings instead of sets containing infinite length strings. To overcome this difficulty, we allow "non-uniform" gales and introduce a new probabilistic notion of *success* of a gale over a distribution. We use this to define two notions of *dimension* of a distribution - strong one and a weak one. Both the notions were already there for the case of infinite strings [25]. In [16], Hitchcock showed that the definition of dimension given by Lutz is equivalent to logarithmic loss unpredictability. In this paper, we show that this result can be adapted to establish a quantitative equivalence between the notion of logarithmic loss unpredictability of a distribution and our proposed notion of dimension. Roughly, this captures the essence of equivalence between pseudorandomness defined via indistinguishability and via unpredictability [36]. We show some important properties of the notion of dimension of a distribution, which eventually makes this characterization much more powerful and flexible. We also do a comparative study between our notion of dimension and two known notions of pseudoentropy, namely HILL-type pseudo min-entropy and next-bit pseudo Shannon entropy. We show that the class of distributions with high dimension is a strict superset of the class of distributions having high HILL-type pseudo min-entropy. Whereas, there is a much closer relationship between dimension and next-bit pseudo Shannon entropy.

Once we have a quantification of pseudorandomness of a distribution, the next natural question is how to extract the pseudorandom part from a given distribution. The question is similar to the question of constructing *randomness extractors* which is an *efficient* algorithm that converts a realistic source to an *almost* ideal source of randomness. The term *randomness extractor* was first defined by Nisan and Zuckerman [28]. Unfortunately there is no such deterministic algorithm and to extract out almost all the randomness, extra $\Omega(\log n)$ pure random bits are always

required [29, 30]. There is a long line of research on construction of extractors towards achieving this bound. For a comprehensive treatment on this topic, we refer the reader to excellent surveys by Nisan and Ta-Shma [26] and Shaltiel [32]. Finally, the desired bound was achieved up to some constant factor in [21].

Coming back to the computational analogue, it is natural to study the same question in the domain of pseudorandomness. Given a distribution with dimension s, the problem is to output $O(sn)$ many bits that are pseudorandom. A simple argument can show that deterministic pseudorandom extraction is not possible, but it is not at all clear that how many pure random bits are necessary to serve the purpose. In this paper, we show that we need to actually involve $\Omega(\log n)$ random bits to extract out all the pseudorandomness present in a distribution. However explicit construction of one such extractor with $O(\log n)$ random bits is not known. If it is known that the given distribution has high HILL-type pseudo min-entropy, then any randomness extractor will work [4]. Instead of HILL-type pseudoentropy, even if we have Yao-type pseudo min-entropy, then also some special kind of randomness extractor (namely with a "reconstruction procedure") could serve our purpose [4]. Unfortunately both of these notions of pseudoentropy can be very small for a distribution with very high dimension. Actually the same counterexample will work for both cases. So it is interesting to come up with an pseudorandom extractor for a class of distributions having high dimension.

As a first step towards this goal, we consider a special kind of source which we call the *nonpseudorandom bit-fixing source*. It is similar to the well studied notion of *bit-fixing random source* introduced by Chor *et al.* [6], for which we know the construction of a deterministic randomness extractor due to [20] and [9]. In this paper, we show that the same construction yields a deterministic pseudorandom extractor for all nonpseudorandom bit-fixing sources having *polynomial-size support*.

In the concluding section, we make a little progress towards the question of P vs. BPP by showing that in order to prove P=BPP, it is sufficient to construct an algorithm that stretches $O(\log n)$ pure random bits to n bits such that the output distribution has a non-zero weak dimension (not necessarily pseudorandom). The idea is that using such stretching algorithm, we easily construct a hard function, which eventually gives us the most desired optimal pseudorandom generator.

Notations: In this paper, we consider the binary alphabet $\Sigma = \{0,1\}$. We denote $Pr_{x\in_R D}[E]$ as $D[E]$, where E is an event and x is drawn randomly according to the distribution D. We use U_m to denote the uniform distribution on Σ^m. Given a string $x \in \Sigma^n$, $x[i]$ denote the i-th bit of x and $x[1,\ldots,i]$ denotes the first i bits of x. Now suppose $x \in \Sigma^n$ and $S = \{s_1, s_2, \ldots, s_k\} \subseteq \{1,2,\ldots,n\}$, then by x_S, we denote the string $x[s_1]x[s_2]\ldots x[s_k]$.

2. Quantification of Pseudorandomness

In this section, we propose a quantification of pseudorandomness present in a distribution. We adapt the notion introduced by Lutz [24] of an s-gale to define a variant notion of success of an s-gale against a distribution D on Σ^n. Throughout this paper, we will talk about non-uniform definitions. First, we consider the definition of pseudorandomness.

2.1. Pseudorandomness

We start by defining the notion of *indistinguishability* which we will use frequently in this paper.

Definition 2.1 (Indistinguishability). A distribution D over Σ^n is (S,ε)-*indistinguishable* from another distribution D' over Σ^n (for $S \in \mathbb{N}, \varepsilon > 0$) if for every circuit C of size at most S,

$$|D[C(x) = 1] - D'[C(x) = 1]| \leq \varepsilon.$$

Now we are ready to introduce the notion of pseudorandomness.

Definition 2.2 (Pseudorandomness). For an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where the distribution D_n is on Σ^n and for any $S = S(n) > n$,[1] $\varepsilon = \varepsilon(n) > 0$,

[1] Throughout this paper, we consider $S(n) > n$ so that the circuit can at least read the full input.

1. (via computational indistinguishability) D is said to be (S,ε)-*pseudorandom* if for all sufficiently large n, D_n is $(O(S(n)),\varepsilon(n))$-indistinguishable from U_n; or equivalently,
2. (via unpredictability [36]) D is said to be (S,ε)-pseudorandom if for all sufficiently large n,

$$D_n[C(x_1,\cdots,x_{i-1})=x_i] \leq \frac{1}{2}+\frac{\varepsilon(n)}{n}$$

for all circuits C of size at most $O(S(n))$ and for all $i \in [n]$.

It is always natural to consider asymptotic definitions with respect to all polynomial size circuits and allowing bias term to be any inverse polynomial. An ensemble of distributions $D=\{D_n\}_{n\in\mathbb{N}}$, where a distribution D_n is on Σ^n, is said to be *pseudorandom* if for every constant $c>0$ and $c'>0$, D is $(n^c, 1/n^{c'})$-pseudorandom [10].

2.2. Martingales, s-gales and predictors

Martingales are "fair" betting games which are used extensively in probability theory (see for example, [3]). Lutz introduced a generalized notion, that of an s-gale, to characterize Hausdorff dimension [23] and Athreya *et al.* used a similar notion to characterize packing dimension[2].

Definition 2.3. [23] Let $s \in [0,\infty)$. An *s-gale* is a function $d:\Sigma^* \to [0,\infty)$ such that $d(\lambda)=1$ and $d(w)=2^{-s}[d(w0)+d(w1)], \forall w \in \Sigma^*$. A *martingale* is a 1-gale.

The following proposition establishes a connection between s-gales and martingales.

Proposition 2.4 ([23]). *A function $d:\Sigma^* \to [0,\infty)$ is an s-gale if and only if the function $d':\Sigma^* \to [0,\infty)$ defined as $d'(w)=2^{(1-s)|w|}d(w)$ is a martingale.*

In order to adapt the notion of an s-gale to the study of pseudorandomness, we first relate it to the notion of predictors, which have been extensively used in the literature [34]. Given an initial finite segment of a string, a predictor specifies a probability distribution over Σ for the next symbol in the string.

Definition 2.5. A function $\pi:\Sigma^*\times\Sigma \to [0,1]$ is a *predictor* if for all $w\in\Sigma^*$, $\pi(w,0)+\pi(w,1)=1$.

Note that the above definition of a predictor is not very different from the type of predictor used in Definition 2.2. If we have a predictor that given a prefix of a string outputs the next bit, then by invoking that predictor independently polynomially many times we can get an estimate on the probability of occurrence of 0 or 1 as the next bit. Using Chernoff bound it can easily be shown that the estimation is correct up to some inverse exponential error. For the detailed equivalence, the reader may refer to [34]. In this paper, we only consider the martingales (or s-gales) and predictors that can be computed using family of non-uniform circuits and from now onwards we refer to them just as martingales (or s-gales) and predictors, and by the size of a martingale (or an s-gale or a predictor), we refer the size of the circuit corresponding to that martingale (or s-gale or predictor).

2.3. Conversion Between s-Gale & Predictor

There is an equivalence between an s-gale and a predictor. An early reference to this is [7]. We follow the construction given in [16].

A predictor π induces an s-gale d_π for each $s\in[0,\infty)$ and is defined as follows: $d_\pi(\lambda)=1$, $d_\pi(wa)=2^s d_\pi(w)\pi(w,a)$ for all $w\in\Sigma^*$ and $a\in\Sigma$; equivalently $d_\pi(w)=2^{s|w|}\prod_{i=1}^{|w|}\pi(w[1\cdots i-1],w[i])$ for all $w\in\Sigma^*$.

Conversely, an s-gale d with $d(\lambda)=1$ induces a predictor π_d defined as: if $d(w)\neq 0$, $\pi_d(w,a)=2^{-s}\frac{d(wa)}{d(w)}$; otherwise, $\pi_d(w,a)=\frac{1}{2}$, for all $w\in\Sigma^*$ and $a\in\Sigma$.

Hitherto, s-gales have been used to study the dimension of sets of infinite sequences - for an extensive bibliography, see [14] and [15]. Although in this paper, we consider distributions on finite length strings, the conversion procedure between s-gale and predictor will be exactly same as described above.

2.4. Defining Dimension

Definition 2.6. For any $\varepsilon > 0$, an s-gale $d : \Sigma^* \to [0,\infty)$ is said to ε-*succeed over a distribution* D_n on Σ^n if

$$D_n[d(w) \geq 2] > \frac{1}{2} + \varepsilon.$$

Note that the above definition of win of an s-gale is not arbitrary and reader may refer to the last portion of the proof of Theorem 4.3 to get some intuition behind this definition. The following lemma states the equivalence between the standard definition of pseudorandomness and the definition using martingale.

Lemma 2.7. *Consider an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where the distribution D_n is on Σ^n. If D is (S,ε)-pseudorandom then for all large enough n, there is no martingale of size at most $O(S(n))$ that $\varepsilon(n)$-succeeds on D_n. Conversely, if for all large enough n, there is no martingale of size at most $O(S(n))$ that $\frac{\varepsilon(n)}{n}$-succeeds on D_n, then D is (S,ε)-pseudorandom.*

Proof. Assume $d : \Sigma^* \to [0,\infty)$ is a martingale which $\varepsilon(n)$-succeeds on D_n for some $n \in \mathbb{N}$, i.e.,

$$D_n[d(w) \geq 2] > \frac{1}{2} + \varepsilon(n).$$

By the Markov Inequality, $U_n[d(w) \geq 2] \leq \frac{1}{2}$.

Let C_d be a circuit of size $O(S(n))$ obtained by instantiating d at length n. Now let C be a circuit which outputs 1 if $C_d(w) \geq 2$. Then,

$$|D_n[C(w) = 1] - U_n[C(w) = 1]| > \varepsilon(n).$$

Thus D is not (S,ε)-pseudorandom.

Now for the converse direction, assume that D is not (S,ε)-pseudorandom. Then there exists an $n_0 \in \mathbb{N}$ such that for any $n \geq n_0$ there exists an bit position $i \in [0,n-1)$ and some circuit C of size at most $O(S(n))$ for which

$$D_n[C(w_1,\cdots,w_{i-1}) = w_i] > \frac{1}{2} + \frac{\varepsilon(n)}{n}.$$

Now build a martingale $d : \Sigma^* \to [0,\infty)$ using this circuit C as follows. Let $d(\lambda) = 1$. Now, $\forall j \in [n], j \neq i$, $d(w[0\ldots j-1]0) = d(w[0\ldots j-1]1) = d(w[0\ldots j-1])$, and $d(w[0\ldots i-1]b) = 2d(w[0\ldots i-1])$, $d(w[0\ldots i-1]\bar{b}) = 0$ if $C(w[0\ldots i-1]) = b$.

Now it is clear that

$$D_n[d(w) \geq 2] > \frac{1}{2} + \frac{\varepsilon(n)}{n}$$

and for all large enough n, the size of the martingale d is at most $O(S(n))$. □

The next definition gives a complete quantification of distributions in terms of dimension.

Definition 2.8 (Weak Dimension or Dimension). For an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where the distribution D_n is on Σ^n and for any $S = S(n) > n$, $\varepsilon = \varepsilon(n) > 0$, the (S,ε)-*weak dimension* or simply *dimension* of D is defined as

$$\dim_{S,\varepsilon}(D) = \inf\{s \in [0,\infty) \mid \text{ for infinitely many } n, \exists s- \text{ gale } d \text{ of size at most } O(S(n)) \text{ which } \varepsilon(n)\text{-succeeds on } D_n\}.$$

Informally, if the dimension of an ensemble of distribution is s, we say that it is s-*pseudorandom*. It is clear from Lemma 2.7 that for any (S,ε)-pseudorandom ensemble of distributions D, $\dim_{S,\varepsilon}(D) \geq 1$. In Section 4 we will see that it is actually an equality. We can also define dimension of distributions in slightly stronger sense.

Definition 2.9 (Strong Dimension). For an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where the distribution D_n is on Σ^n and for any $S = S(n) > n$, $\varepsilon = \varepsilon(n) > 0$, the (S,ε)-*strong dimension* of D is defined as

$$\mathrm{sdim}_{S,\varepsilon}(D) = \inf\{s \in [0,\infty) \mid \text{ for all large enough } n, \exists s-\text{ gale } d \text{ of size at most } O(S(n)) \text{ which } \varepsilon(n)\text{-succeeds on } D_n\}.$$

It follows from the definition that weak dimension is smaller or equal to strong dimension.

Proposition 2.10. *For an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where the distribution D_n is on Σ^n and for any $S = S(n) > n$, $\varepsilon = \varepsilon(n) > 0$, $\dim_{S,\varepsilon}(D)$ and $\mathrm{sdim}_{S,\varepsilon}(D)$ are well defined and*

$$\dim_{S,\varepsilon}(D) \leq \mathrm{sdim}_{S,\varepsilon}(D).$$

We will show separation between both of these notions of dimensions in Section 4. Now just like the asymptotic definition of pseudorandomness with respect to all polynomial size circuits and inverse polynomial bias, for any ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, we can give definitions of weak dimension or simply dimension (denoted by $\dim(D)$) and strong dimension (denoted by $\mathrm{sdim}(D)$) by allowing $S(n)$ to be $n^{O(1)}$ and $\varepsilon(n)$ to be $n^{-O(1)}$ in the Definition 2.8 and 2.9 respectively.

3. Unpredictability and Dimension

It is customary to measure the performance of a predictor utilizing a *loss function* [17]. The loss function determines the penalty incurred by a predictor for erring in its prediction. Let the next bit be b and the probability induced by the predictor on it is p_b.

Commonly used loss functions include the *absolute loss function*, which penalizes the amount $1 - p_b$; and the *logarithmic loss function*, which penalizes $-\log(p_b)$. The latter, which appears complicated at first glance, is intimately related to the concepts of Shannon Entropy and dimension. In this section, adapting the result of Hitchcock [16], we establish that there is an equivalence between the notion of dimension that we have defined in the previous section, and the logarithmic loss function defined on a predictor.

Definition 3.1. The *logarithmic loss function* on $p \in [0,1]$ is defined to be $loss(p) = -\log p$.

Using this, we define the running loss that a predictor incurs while it predicts successive bits of a string in Σ^n, as the sum of the losses that the predictor makes on individual bits.

Definition 3.2. Let $\pi : \Sigma^* \times \Sigma \to [0,1]$ be a predictor.

1. The *cumulative loss* of π on $w \in \Sigma^n$, denoted as $\mathrm{Loss}(\pi, w)$, is defined by $\mathrm{Loss}(\pi, w) = \sum_{i=1}^{n} loss(\pi(w[1\ldots i-1],\, w[i]))$.
2. The *loss rate* of π on $w \in \Sigma^n$ is $\mathrm{LossRate}(\pi, w) = \frac{\mathrm{Loss}(\pi,w)}{n}$.
3. The ε-*loss rate* of π over a distribution D_n on Σ^n is

$$\mathrm{LossRate}_\varepsilon(\pi, D_n) = \inf\{t \in [0,1] \mid D_n[\mathrm{LossRate}(\pi, w) \leq t] > \frac{1}{2} + \varepsilon\}.$$

Intuitively the unpredictability of a distribution is defined as the infimum of the loss rate that any predictor has to incur on the distribution.

Definition 3.3 (Weak Unpredictability or Unpredictability). For an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where the distribution D_n is on Σ^n and for any $S = S(n) > n$, $\varepsilon = \varepsilon(n) > 0$, the (S,ε)-*weak unpredictability* or simply *unpredictability* of D is

$$\mathrm{unpred}_{S,\varepsilon}(D) = \inf\{t \in [0,1] \mid \text{ for infinitely many } n, \text{ there exists a predictor } \pi \text{ of size at most } O(S(n)) \text{ such that } \mathrm{LossRate}_{\varepsilon(n)}(\pi, D_n) \leq t\}.$$

With this, we can prove that dimension can equivalently be defined using unpredictability. The proof is motivated from the proof of the equivalence between logarithmic loss unpredictability and dimension [16].

Theorem 3.4. *Consider any $s \in [0,1]$. For an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where the distribution D_n is on Σ^n and for any $S = S(n) > n$, $\varepsilon = \varepsilon(n) > 0$, if $\dim_{S,\varepsilon}(D) \leq s$ then $\mathrm{unpred}_{S^{O(1)},\varepsilon}(D) \leq s$. Conversely, $\mathrm{unpred}_{S,\varepsilon}(D) \leq s$ implies $\dim_{S^{O(1)},\varepsilon}(D) \leq s$.*

Proof. Assume that s' is any number such that $s < s'$ and then take a number s'' such that $s < s'' \leq s'$ and $2^{s''}$ is rational[2]. For some large enough n, suppose d is an s''-gale of size at most $O(S(n))$ that $\varepsilon(n)$-succeeds on D_n. Let $\pi_d : \Sigma^* \times \Sigma \to [0,1]$ be defined by

$$\pi_d(w,b) = \begin{cases} 2^{-s''}\frac{d(wb)}{d(w)} & \text{if } d(w) \neq 0 \\ \frac{1}{2} & \textit{otherwise.} \end{cases}$$

For any $w \in \Sigma^n$ with $d(w) \geq 2$, we have

$$\begin{aligned} \mathrm{Loss}_{\pi_d}(w) &= -\sum_{i=1}^{n} \log \pi_d(w[1\ldots i-1], w[i]) \\ &= -\log \Pi_{i=1}^{n} \pi_d(w[1\ldots i-1], w[i]) \\ &= s''n - \log d(w) \\ &\leq s''n - 1 \leq s'n. \end{aligned}$$

So $\mathrm{LossRate}(\pi_d, w) \leq s'$. Thus,

$$D_n[\mathrm{LossRate}(\pi_d, w) \leq s'] \geq D_n[d(w) \geq 2] > \frac{1}{2} + \varepsilon.$$

Note that implementation of π_d involves division of two at most $O(S(n))$ bits rational numbers[3] and thus can be done using a circuit of size at most $(S(n))^{O(1)}$ [35].

Conversely, assume that $\mathrm{unpred}_{S,\varepsilon}(D) \leq t \in [0,1]$. Assume that t' is any number satisfying $t < t'$ and then take any number t'' such that $t' < t''$ and $2^{t''}$ is rational. For some large enough n, let π be a predictor of size at most $O(S(n))$ such that

$$D_n[\mathrm{LossRate}(\pi, w) \leq t'] > \frac{1}{2} + \varepsilon.$$

If d_π is the t''-gale defined by

$$d_\pi(w) = 2^{t''|w|} \Pi_{i=1}^{|w|} \pi(w[0\ldots i-1], w[i])$$

then for any $w \in \Sigma^n$ with $\mathrm{LossRate}(\pi, w) \leq t'$, we have the following,

$$\begin{aligned} \log d_\pi(w) &= t''n + \sum_{i=1}^{n} \log \pi(w[1\ldots i-1], w[i]) \\ &= t''n - \mathrm{Loss}_\pi(w) \geq 1. \end{aligned}$$

[2]We consider $2^{s''}$ to be rational to ensure that the value of $2^{s''}$ can be computed using constant size circuit. Note that we can always find such s'' due to the fact that the function 2^s for $s > 0$ is continuous, monotonically increasing and within any two real numbers there exists a rational number.

[3]As we are considering martingales and predictors that can be implemented by circuits of size $O(S(n))$ so the output must be a rational number which can be represented by at most $O(S(n))$ bits.

The last inequality holds for all sufficiently large n. Hence, for infinitely many n,

$$D_n[d_\pi(w) \geq 2] > \frac{1}{2} + \varepsilon.$$

Moreover, computation of d involves multiplication of n rational numbers of at most $O(S(n))$ bits each and thus can be implemented by a circuit of size $(S(n))^{O(1)}$ [35]. □

Just like dimension, one can also define strong unpredictability in the following way.

Definition 3.5 (Strong Unpredictability). For an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where the distribution D_n is on Σ^n and for any $S = S(n) > n$, $\varepsilon = \varepsilon(n) > 0$, the (S,ε)-*strong unpredictability* of D is

$$\text{sunpred}_{S,\varepsilon}(D) = \inf\{t \in [0,1] \mid \text{ for all large enough } n, \text{ there exists a predictor } \pi \text{ of size at most } O(S(n)) \text{ such that } \text{LossRate}_{\varepsilon(n)}(\pi, D_n) \leq t\}.$$

Now by following the proof of Theorem 3.4 it is easy to see that a similar relation also holds between strong dimension and strong unpredictability.

Theorem 3.6. *Consider any $s \in [0,1]$. For an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where the distribution D_n is on Σ^n and for any $S = S(n) > n$, $\varepsilon = \varepsilon(n) > 0$, if $\text{sdim}_{S,\varepsilon}(D) \leq s$ then $\text{sunpred}_{S^{O(1)},\varepsilon}(D) \leq s$. Conversely, $\text{sunpred}_{S,\varepsilon}(D) \leq s$ implies $\text{sdim}_{S^{O(1)},\varepsilon}(D) \leq s$.*

Analogous to pseudorandomness and dimension, for any ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, one can define weak unpredictability or simply unpredictability (denoted by $\text{unpred}(D)$) and strong unpredictability (denoted by $\text{sunpred}(D)$) by allowing $S(n)$ to be $n^{O(1)}$ and $\varepsilon(n)$ to be $n^{-O(1)}$ in the Definition 3.3 and 3.5 respectively. Following is a straight forward implication of Theorem 3.4 and Theorem 3.6.

Corollary 3.7. *For any ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where the distribution D_n is on Σ^n,*

$$\dim(D) = \text{unpred}(D) \text{ and } \text{sdim}(D) = \text{sunpred}(D).$$

4. Properties of Dimension

We now establish a few basic properties of our notion of dimension. We begin by exhibiting existence of an ensemble of distributions with dimension s, for any $s \in [0,1]$.

First, we observe that the dimension of any ensemble of distributions D is the infimum of a non-empty subset of $[0, 1+\varepsilon]$ for any $\varepsilon > 0$ and hence the dimension of a distribution is well-defined. The following lemma establishes the above claim.

Lemma 4.1. *For an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where the distribution D_n is on Σ^n and for any $S = S(n) > n$, $\varepsilon = \varepsilon(n) > 0$, $\text{sdim}_{S,\varepsilon}(D) \leq 1$.*

Proof. Let us first take any $s > 1$ such that 2^s is rational and then consider the following function $d : \Sigma^* \to [0,\infty)$. Let $d(\lambda) = 1$ and $\forall i \in [n]$, $d(w[0\ldots i-1]0) = d(w[0\ldots i-1]1) = 2^{s-1}d(w[0\ldots i-1])$. It is easy to see that d is an s-gale and for any $w \in \Sigma^n$, $d(w) = 2^{(s-1)n}$. Thus for all large enough n, d will $\varepsilon(n)$-succeed over D_n. Also note that for all large enough n, this function d can be implemented using a circuit of size $O(n)$.[4] Hence the statement of the lemma follows. □

Above lemma along with Proposition 2.10 implies the following corollary.

[4] As for every string w of length n, the value of $d(w)$ will be same, so one can hardcode the value $2^{(s-1)n}$ inside the non-uniform circuit implementing the function d.

Corollary 4.2. *For an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where the distribution D_n is on Σ^n and for any $S = S(n) > n$, $\varepsilon = \varepsilon(n) > 0$, $\dim_{S,\varepsilon}(D) \leq 1$.*

Since it is clear that any ensemble of distributions has a dimension, the following theorem establishes the fact that our definition yields a nontrivial quantification of the set of ensembles of distributions.

Theorem 4.3. *Let $s \in [0,1]$. Then for for any $S = S(n) > n$, $\varepsilon = \varepsilon(n) > 0$, there is an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where the distribution D_n is on Σ^n, such that (S,ε)-dimension (also strong dimension) of D is s.*

Proof. Let us take the ensemble of uniform distributions, i.e., $D := \{U_n\}_{n\in\mathbb{N}}$, where U_n is the uniform distribution on Σ^n. Then by Lemma 2.7 together with Lemma 4.1 shows that for any S and ε, $\mathrm{sdim}_{S,\varepsilon}(D) = 1$. By using corollary 4.2 instead of Lemma 4.1, one can also show that $\dim_{S,\varepsilon}(D) = 1$.

On the other hand we consider an ensemble $D = \{D_n\}_{n\in\mathbb{N}}$ where support size of each D_n is one, or in other words, D_n imposes all the probability on a single string, say 0^n. Now first take any $s > 0$ such that 2^s is rational and then consider the following function $d : \Sigma^* \to [0,\infty)$. Let $d(\lambda) = 1$ and $\forall i \in [n]$, $d(w[0\ldots i-1]0) = 2^s d(w[0\ldots i-1])$. It is easy to see that d is an s-gale and $d(0^n) = 2^{sn}$. Thus for all large enough n, d will $\varepsilon(n)$-succeed over D_n. Also note that for all large enough n this function d can be implemented using a circuit of size $O(n)$. Hence for any S and ε, (S,ε)-dimension as well as (S,ε)-strong dimension of D is 0.

Otherwise, assume that $s \in (0,1)$. Let us take the ensemble of uniform distributions $D := \{U_n\}_{n\in\mathbb{N}}$. To each string $x \in \Sigma^n$, we append $\lfloor \frac{n}{s} \rfloor - n$ many zeros, and denote the resulting string as x'. Let $D'_n(x') = U_n(x)$. For strings $y \in \Sigma^{\lfloor \frac{n}{s} \rfloor}$ which do not terminate in a sequence of $\lfloor \frac{n}{s} \rfloor - n$ many zeros, we set $D'_n(y) = 0$.

For any large enough n, let $\pi : \Sigma^* \times \Sigma \to [0,1]$ be the predictor for distribution U_n which testifies that the $(S^{O(1)},\varepsilon)$-strong unpredictability of D is at most 1. Define the new predictor $\pi' : \Sigma^* \times \Sigma \to [0,1]$ by

$$\pi'(x,b) = \begin{cases} \pi(x,b) & \text{if} |x| < n, b = 0,1 \\ 1 & \text{if} |x| \geq n, b = 0 \\ 0 & \text{otherwise.} \end{cases}$$

For every $w \in \Sigma^{\lfloor \frac{n}{s} \rfloor}$ which is in the support of D'_n such that $\mathrm{LossRate}(\pi, w[1\ldots n]) \leq (1+\varepsilon_1)$, for any $\varepsilon_1 > 0$, we have that

$$\mathrm{LossRate}(\pi', w) = \frac{\mathrm{Loss}(\pi, w[1\ldots n])}{\lfloor \frac{n}{s} \rfloor} \leq \frac{(1+\varepsilon_1)n}{\lfloor \frac{n}{s} \rfloor} \leq (s+\varepsilon'), \text{ for some } \varepsilon' > 0$$

The last inequality holds for all small enough s/n and this testifies that the $(S^{O(1)},\varepsilon)$-strong unpredictability (hence the $(S^{O(1)},\varepsilon)$-strong dimension) of the ensemble of distributions D' is at most s. Now by Proposition 2.10, it follows that $\dim_{S^{O(1)},\varepsilon}(D')$ is also at most s.

Now, assume that (S,ε)-dimension of D' is less than s. For any s' such that $0 < s' < s$, for infinitely many n, there exists an s'-gale d of size at most $O(S(n))$ which $\varepsilon(n)$-succeeds on D'_n. We show that this would imply that U_n is not uniform. Now consider a string $w \in \Sigma^{\lfloor \frac{n}{s} \rfloor}$, which is in the support of D'_n. For any $k \in \{n+1, \cdots, \lfloor \frac{n}{s} \rfloor\}$, $d(w[1\ldots k]) \leq 2^{s'} d(w[1\ldots k-1])$ and thus $d(w) \geq 2$ will imply that $d(w[1\ldots n]) \geq 2^{-s'(\lfloor \frac{n}{s} \rfloor - n)+1}$. Now consider the martingale d' (needs not be computed by any circuit) corresponding to the s'-gale d. According to [23], we have $d'(w') = 2^{(1-s')|w'|} d(w')$, for any string $w' \in \Sigma^*$. Thus,

$$\begin{aligned} D'_n[d'(w[1\ldots n]) \geq 2] &\geq D'_n[d(w[1\ldots n]) \geq 2^{-s'(\lfloor \frac{n}{s} \rfloor - n)+1}] \\ &\geq D'_n[d(w) \geq 2] \\ &> \frac{1}{2} + \varepsilon(n). \end{aligned}$$

Note that $D'_n[d'(w[1\dots n]) \geq 2]$ is same as $U_n[d'(x) \geq 2]$, which contradicts the fact that by Markov inequality, $U_n[d'(x) \geq 2] \leq \frac{1}{2}$. This shows that $\dim_{S,\varepsilon}(D') \geq s$ and hence by Proposition 2.10, $\mathrm{sdim}_{S,\varepsilon}(D')$ is also greater than or equal to s. □

Informally the following theorem shows that our notion of dimension is able to capture the fact that if we mix a "good" distribution with a small amount of an extremely "bad" distribution, then also "quality" of the first distribution would not change much.

Theorem 4.4. *Let $D = \{D_n\}_{n\in\mathbb{N}}$, $D^1 = \{D^1_n\}_{n\in\mathbb{N}}$ and $D^2 = \{D^2_n\}_{n\in\mathbb{N}}$ be three ensembles of distributions such that for some $\delta = \delta(n) \in [0,1]$, for every $n \in \mathbb{N}$, $D_n = (1-\delta(n))D^1_n + \delta(n)D^2_n$. If for any $S = S(n) > n$ and $\varepsilon = \varepsilon(n) > 0$, $\dim_{S,\varepsilon}(D^1) = s_1$ ($\mathrm{sdim}_{S,\varepsilon}(D^1) = s_1$), then $\dim_{S,(\varepsilon+\delta)}(D) \geq s_1$ ($\mathrm{sdim}_{S,(\varepsilon+\delta)}(D) \geq s_1$).*[5]

Proof. For the contrary, let us assume that, $\dim_{S,(\varepsilon+\delta)}(D) < s_1$ and assume $s = s_1 - \varepsilon_1$, for some ε_1, $0 < \varepsilon_1 < s_1$. So for infinitely many n, there exists an s-gale of size at most $O(S(n))$ that $(\varepsilon(n)+\delta(n))$-succeeds over D_n. Thus,

$$D_n[d(w) \geq 2] > \frac{1}{2} + (\varepsilon(n) + \delta(n)).$$

Let the strings w for which $d(w) \geq 2$ holds be w_i, $1 \leq i \leq k$. So,

$$D_n(w_1) + \cdots + D_n(w_k) > \frac{1}{2} + (\varepsilon(n) + \delta(n)),$$

where $D_n(w_i) = (1-\delta(n))D^1_n(w_i) + \delta(n)D^2_n(w_i)$, for all $1 \leq i \leq k$.
Now, as $D^2_n(w_1) + \cdots + D^2_n(w_k) \leq 1$,

$$D^1_n(w_1) + \cdots + D^1_n(w_k) > \frac{1}{2} + \varepsilon(n)$$

and thus $\dim_{S,\varepsilon}(D^1) < s_1$ which is a contradiction.

The above argument is valid for all n for which there exists an s-gale of size at most $O(S(n))$ that $(\varepsilon(n)+\delta(n))$-succeeds over D_n and hence the same claim holds even if we consider strong dimension instead of dimension. □

If we follow the proof of Theorem 4.4 with martingale instead of s-gale, we get the following weaker version of the above theorem, which we will require in the construction of deterministic extractor for a special kind of sources in Section 6.1.

Lemma 4.5. *Let $D = \{D_n\}_{n\in\mathbb{N}}$, $D^1 = \{D^1_n\}_{n\in\mathbb{N}}$ and $D^2 = \{D^2_n\}_{n\in\mathbb{N}}$ be three ensembles of distributions such that for some $\delta = \delta(n) \in [0,1]$, for every $n \in \mathbb{N}$, $D_n = (1-\delta(n))D^1_n + \delta(n)D^2_n$. If for any $S = S(n) > n$ and $\varepsilon = \varepsilon(n) > 0$, D^1 is (S,ε)-pseudorandom, then D is $(S,\varepsilon+\delta)$-pseudorandom.*

In subsequent sections, we will see how to extract pseudorandom parts from a convex combination of distributions. For that purpose we need the following lemma which establishes the fact that convex combination of two pseudorandom distributions will be pseudorandom as well.

Lemma 4.6. *Consider two ensembles of distributions $D^1 = \{D^1_n\}_{n\in\mathbb{N}}$ and $D^2 = \{D^2_n\}_{n\in\mathbb{N}}$. Suppose for any $S = S(n) > n$ and $\varepsilon = \varepsilon(n) > 0$, both D^1 and D^2 are (S,ε)-pseudorandom. If for $\delta = \delta(n) \in [0,1]$ there exists an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$ which can be expressed as for all $n \in \mathbb{N}$, $D_n = \delta(n)D^1_n + (1-\delta(n))D^2_n$, then D is also (S,ε')-pseudorandom, where $\varepsilon'(n) = n \cdot \varepsilon(n)$.*[6]

[5] Note that bias term in the dimension of D^1 depends on δ.

[6] Note that this lemma will be useful only when we consider $\varepsilon(n) < \frac{1}{2n}$.

Proof. The claim clearly holds when $\delta(n)$ is either 0 or 1, so assume that $0 < \delta(n) < 1$. For the contrary, let us assume that D is not (S,ε')-pseudorandom where $\varepsilon'(n) = n \cdot \varepsilon(n)$, i.e., by Lemma 2.7, for infinitely many $n \in \mathbb{N}$ there exists an martingale d of size at most $O(S(n))$ that $\varepsilon(n)$-succeeds on D_n, i.e.,

$$D_n[d(w) \geq 2] > \frac{1}{2} + \varepsilon(n).$$

As D^2 is (S,ε)-pseudorandom, by Lemma 2.7, for all large enough n there exists no martingale of size at most $O(S(n))$ that $\varepsilon(n)$-succeeds on D_n^2. Thus it is possible to consider an $n \in \mathbb{N}$ such that there exists a martingale d of size at most $O(S(n))$ that $\varepsilon(n)$-succeeds on D_n, but does not $\varepsilon(n)$-succeed on D_n^2. Let the strings $w \in \Sigma^n$ for which $d(w) \geq 2$ holds be w_i, $1 \leq i \leq k$. So,

$$D_n(w_1) + \cdots + D_n(w_k) > \frac{1}{2} + \varepsilon(n),$$

where $D_n(w_i) = \delta(n)D_n^1(w_i) + (1-\delta(n))D_n^2(w_i), 1 \leq i \leq k$. Now, since we have that

$$D_n^2(w_1) + \cdots + D_n^2(w_k) \leq \frac{1}{2} + \varepsilon(n).$$

Thus the following holds

$$D_n^1(w_1) + \cdots + D_n^1(w_k) > \frac{1}{2} + \varepsilon(n).$$

As for infinitely many $n \in \mathbb{N}$ the above happens, so D^1 is not (S,ε)-pseudorandom, which is a contradiction. □

We can also slightly generalize the above lemma and get the following theorem.

Theorem 4.7. *Let $D = \{D_n\}_{n\in\mathbb{N}}$, $D^1 = \{D_n^1\}_{n\in\mathbb{N}}$ and $D^2 = \{D_n^2\}_{n\in\mathbb{N}}$ be three ensembles of distributions such that for some $\delta = \delta(n) \in [0,1]$, for every $n \in \mathbb{N}$, $D_n = \delta(n)D_n^1 + (1-\delta(n))D_n^2$. Then for any $S = S(n) > n$ and $\varepsilon = \varepsilon(n) > 0$,*

$$\dim_{S,\varepsilon}(D) \geq \min\{\dim_{S,\varepsilon}(D^1), \dim_{S,\varepsilon}(D^2)\}.$$

Proof. The claim clearly holds when $\delta(n)$ is either 0 or 1, so assume that $0 < \delta(n) < 1$. Let $\dim_{S,\varepsilon}(D_1) = s_1$, and $\dim_{S,\varepsilon}(D_2) = s_2$.

For the contrary, let us assume that, $\dim_{S,\varepsilon}(D) < \min\{s_1, s_2\}$. Now consider $s = \min\{s_1, s_2\} - \varepsilon_1$, for some $\varepsilon_1, 0 < \varepsilon_1 < \min\{s_1, s_2\}$. Then for infinitely many n, there exists an s-gale d of size at most $O(S(n))$ such that

$$D_n[d(w) \geq 2] > \frac{1}{2} + \varepsilon(n).$$

As $\dim_{S,\varepsilon}(D^2) = s_2$, so for all large enough n, there exists no s-gale of size at most $O(S(n))$ that $\varepsilon(n)$-succeeds on D_n^2. Thus we can consider an n such that there exists an s-gale d of size at most $O(S(n))$ that $\varepsilon(n)$-succeeds on D_n, but does not $\varepsilon(n)$-succeed on D_n^2. Let the strings w for which $d(w) \geq 2$ holds be w_i, $1 \leq i \leq k$. So,

$$D_n(w_1) + \cdots + D_n(w_k) > \frac{1}{2} + \varepsilon(n)$$

where $D_n(w_i) = \delta(n)D_n^1(w_i) + (1-\delta(n))D_n^2(w_i)$, for $1 \leq i \leq k$. Also, we have that

$$D_n^2(w_1) + \cdots + D_n^2(w_k) \leq \frac{1}{2} + \varepsilon(n).$$

Thus

$$D_n^1(w_1) + \cdots + D_n^1(w_k) > \frac{1}{2} + \varepsilon(n)$$

and this happens for infinitely many $n \in \mathbb{N}$ implying $\dim_{S,\varepsilon}(D^1) < s_1$, which is a contradiction. □

The following theorem shows that in order for a distribution to have dimension less than 1, it is not sufficient to have a few positions where we can successfully predict - it is necessary that these positions occur often.

Theorem 4.8. *For any $S = S(n) > n$ and $\varepsilon = \varepsilon(n) > 0$, there is an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$ such that $\dim_{S,\varepsilon}(D) = 1$, but D is not (S,ε)-pseudorandom.*

Proof. Let D_n on Σ^n be defined as follows.

$$D_n(x) = \begin{cases} \frac{1}{2^{n-1}} & \text{if } x[n] = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then D is clearly not (S,ε)-pseudorandom, for any value of $S = S(n) > n$ and $\varepsilon = \varepsilon(n) > 0$. Consider a predictor $\pi : \Sigma^* \times \Sigma \to [0,1]$ defined as follows. For strings w of length i, $i \in [1, n-2]$, set $\pi(w,b) = 0.5$, $b = 0,1$ and $\pi(w,0) = 1$, $\pi(w,1) = 0$ otherwise. Then

$$D_n[\pi(x[\ 1 \ldots n-1\ ], x[n]) = 1] = 1.$$

However, we will show that $\dim_{S,\varepsilon}(D) = 1$. Assume that $\dim_{S,\varepsilon}(D) < 1$ and thus for some ε_1, $0 < \varepsilon_1 < 1$, for infinitely many $n \in \mathbb{N}$, there exists an s-gale d, where $s = 1 - \varepsilon_1$, of size at most $O(S(n))$ which $\varepsilon(n)$-succeeds on D_n. Now consider a string $w \in \Sigma^n$, which is in the support of D_n. Now, $d(w) \leq 2^s d(w[1 \ldots n-1])$ and thus $d(w) \geq 2$ will imply that $d(w[1 \ldots n-1]) \geq 2^{1-s}$. Next consider the martingale d' (needs not be computed by any circuit) corresponding to the s-gale d. According to [23], we have $d'(w') = 2^{(1-s)|w'|} d(w')$, for any string $w' \in \Sigma^*$. Thus,

$$\begin{aligned} D_n[d'(w[1 \ldots n-1]) \geq 2] &\geq D_n[d(w[1 \ldots n-1]) \geq 2^{1-s}] \\ &\geq D_n[d(w) \geq 2] \\ &> \frac{1}{2} + \varepsilon(n). \end{aligned}$$

The first inequality holds for all large enough n. Note that $D_n[d'(w[1 \ldots n-1]) \geq 2]$ is same as $U_{n-1}[d'(x) \geq 2]$, where $x \in \Sigma^{n-1}$ is drawn according to the distribution U_{n-1}. However by Markov inequality, $U_{n-1}[d'(x) \geq 2] \leq \frac{1}{2}$, which is a contradiction and this completes the proof. □

We now give separation between two notions of dimension by providing an example of ensemble which has a very high strong dimension, but very low weak dimension.

Theorem 4.9. *There exists an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$ such that for any $S = S(n) > n$ and $\varepsilon = \varepsilon(n) > 0$, $\dim_{S,\varepsilon}(D) < \mathrm{sdim}_{S,\varepsilon}(D)$.*

Proof. Here we actually show a much stronger claim by giving an example of an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$ such that $\mathrm{sdim}_{S,\varepsilon}(D) = 1$ whereas $\dim_{S,\varepsilon}(D) = 0$. This is the largest possible gap between weak and strong dimension of any ensemble of distributions.

Construct an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$ as follows: for all the odd value of n, set $D_n = U_n$ where U_n is the uniform distribution over Σ^n and for all even value of n, set D_n to be such that it imposes all the probability on a single string 0^n.

Due to Markov inequality, there exists no martingale that $\varepsilon(n)$-succeeds over distribution D_n for any odd value of n and by Lemma 4.1, strong dimension can be at most 1. Hence, $\mathrm{sdim}_{S,\varepsilon}(D) = 1$. By the argument used in the

proof of Theorem 4.3, we can say that for any $s > 0$ such that 2^s is rational, for all large enough even value of n, there exists an s-gale of size at most $O(S(n))$ that $\varepsilon(n)$-succeeds over D_n and hence $\dim_{S,\varepsilon}(D) = 0$. □

5. Pseudoentropy and Dimension

In this section we study the relation between our notion of dimension and different variants of computational or pseudo (min/Shannon) entropy. We will use standard notions and notations of information theory (e.g., Shannon entropy, KL divergence) without defining them. Readers can find a few basic notations and propositions of information theory in the following subsection and for more details, we refer the reader to the book by Cover and Thomas [8].

5.1. Basics of Information Theory

Definition 5.1 (Shannon Entropy). The *Shannon entropy* of a discrete random variable X is defined as

$$H(X) := -\sum_x Pr[X = x] \log Pr[X = x] = -\mathbb{E}_{x\sim X}[\log Pr[X = x]].$$

The *joint entropy* $H(X,Y)$ is defined to be $-\mathbb{E}_{x\sim X, y\sim Y}[\log Pr[X = x, Y = y]]$ and the *conditional entropy* $H(Y \mid X)$ is defined to be $\mathbb{E}_{x\sim X}[H(Y \mid X = x)]$.

Proposition 5.2 (Chain Rule for Shannon Entropy).

$$H(X,Y) = H(X) + H(Y \mid X).$$

Definition 5.3 (KL divergence). The *Kullback-Leibler distance* or *KL divergence* between two distributions P and Q is defined as

$$\mathbf{KL}(P\|Q) := \mathbb{E}_{p\sim P} \log \frac{Pr[P = p]}{Pr[Q = p]}.$$

Definition 5.4 (Conditional KL divergence). For random variables (P_1, P_2) and (Q_1, Q_2), the *conditional KL divergence* from $(P_2|P_1)$ to $(Q_2|Q_1)$ is defined as

$$\mathbf{KL}((P_2|P_1)\|(Q_2|Q_1)) = \mathbb{E}_{p_1\sim P_1, p_2\sim P_2}\big[\log \frac{Pr[P_2 = p_2|P_1 = p_1]}{Pr[Q_2 = p_2|Q_1 = p_1]}\big].$$

Just like Shannon entropy, in this case also, we have chain rule stated below.

Proposition 5.5 (Chain Rule for KL divergence).

$$\mathbf{KL}(P_1, P_2\|Q_1, Q_2) = \mathbf{KL}(P_1\|Q_1) + \mathbf{KL}((P_2|P_1)\|(Q_2|Q_1)).$$

5.2. High HILL-type pseudo min-entropy implies high dimension

For a distribution D, the *min-entropy* of D is defined as $H_\infty(D) = \min_w\{\log(1/D[w])\}$. We start with the standard definition of computational min-entropy, as given by [13].

Definition 5.6 (HILL-type pseudo min-entropy). For any $S = S(n) > n$ and $\varepsilon = \varepsilon(n) > 0$, an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$ has (S,ε)-*HILL-type pseudo min-entropy* (or simply (S,ε)-pseudo min-entropy) at least $k = k(n)$, denoted as $H_\infty^{HILL,S,\varepsilon}(D) \geq k$ if there exists an ensemble of distributions $D' = \{D'_n\}_{n\in\mathbb{N}}$ such that for all large enough n,

1. $H_\infty(D'_n) \geq k(n)$, and
2. D'_n is $(O(S(n)), \varepsilon(n))$-indistinguishable from the distribution D_n.

Several other definitions of pseudo min-entropy (metric-type, Yao-type or compression type) are present in the literature. We refer the reader to [4] for a comprehensive treatment on different definitions and the connections between them. In the remaining portion of this subsection, we focus only on HILL-type pseudo min-entropy. Now we state the main result of this subsection.

Theorem 5.7. *For every ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$ and for any $S = S(n) > n$, $\varepsilon = \varepsilon(n) > 0$, if $H_\infty^{HILL,S,\varepsilon}(D) \geq k$, where $k = k(n) = sn$ for some $s \in [0,1]$, then $\dim_{S,\varepsilon}(D) \geq s$.*

Proof. The theorem is a consequence of the following lemma.

Lemma 5.8. *For every ensemble of distributions $X = \{X_n\}_{n\in\mathbb{N}}$, if for all large enough n, $H_\infty(X_n) \geq k(n)$, where $k(n) = sn$ for some $s \in [0,1]$, then $\dim_{S,\varepsilon}(X) \geq s$ for any $S = S(n) > n$ and $\varepsilon = \varepsilon(n) > 0$.*

Now observe that if for all large enough n, distribution D_n is $(O(S(n)), \varepsilon(n))$-indistinguishable from another distribution D'_n, then $\dim_{S,\varepsilon}(D) = \dim_{S,\varepsilon}(D')$ as otherwise the s-gale of size at most $O(S(n))$ which $\varepsilon(n)$-succeeds over exactly one of them, acts as a distinguishing circuit. This fact along with Lemma 5.8 completes the proof. □

It only remains to establish Lemma 5.8.

Proof of Lemma 5.8. For the sake of contradiction, let us assume that for infinitely many $n \in \mathbb{N}$, there exists an s-gale d of size at most $O(S(n))$ that $\varepsilon(n)$-succeeds over X_n, i.e.,

$$X_n[d(w) \geq 2] > \frac{1}{2} + \varepsilon(n).$$

Now consider the following set

$$S := \{w \in \Sigma^n \mid d(w) \geq 2\}.$$

As $H_\infty(X_n) \geq k(n)$,

$$|S| > 2^{sn-1} + 2^{sn} \cdot \varepsilon(n).$$

By taking the corresponding martingale d' (needs not be computed by any circuit) according to the Proposition 2.4, we have that for any $w \in S$, $d'(w) \geq 2^{(1-s)n+1}$ and as a consequence,

$$U_n[d'(w) \geq 2^{(1-s)n+1}] > 2^{sn-n-1} + 2^{sn-n} \cdot \varepsilon(n)$$

which contradicts the fact that by Markov inequality, $U_n[d'(w) \geq 2^{(1-s)n+1}] \leq 2^{sn-n-1}$. □

One can extend the definition of HILL-type pseudo min-entropy by allowing $S(n)$ to be $n^{O(1)}$ and $\varepsilon(n)$ to be $n^{-O(1)}$ and let us denote this by $H_\infty^{HILL}(D)$. We can extend the result stated in Theorem 5.7 as follows.

Corollary 5.9. *For any ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where D_n is a distribution over Σ^n,and $s \in [0,1]$, if $H_\infty^{HILL}(D_n) \geq k$, where $k = k(n) = sn$, then $\dim(D) \geq s$.*

The converse direction of the statement of Theorem 5.7 is also true if the distribution under consideration is pseudorandom. If the converse is true then we can apply any randomness extractor to get pseudorandom distribution from any distribution having high dimension [4]. However, we should always be careful about the circuit size with respect to which we call the output distribution pseudorandom. Unfortunately, in general the converse is not true.

Counterexample for the converse: Suppose *one-way functions* (see Definition 2.2.1 of [10]) exist, then it is well-known that we can construct a *pseudorandom generator* (see Definition 3.3.1 of [10]) $G = \{G_l\}_{l\in\mathbb{N}}$ where $G_l : \Sigma^l \to \Sigma^m$ such that m is any polynomial in l, say $m = l^3$. For the construction of pseudorandom generator with polynomial stretch from any one-way function, interested reader may refer to [13, 34]. Now consider the ensemble of distributions $D := \{(G(U_l), U_l)\}_{l\in\mathbb{N}}$. For large enough l, using the argument similar to the proof of Theorem 4.3, it can easily be shown that the distribution D has dimension (as well as weak dimension) almost 1 as the distribution on the first m bits are pseudorandom, but pseudo min-entropy is not larger than l.

5.3. Equivalence between dimension and next-bit pseudo Shannon entropy

In the last subsection, we have talked about pseudo min-entropy. In similar fashion, one can also define *pseudo Shannon entropy* and a natural generalization of it is *conditional pseudo Shannon entropy* [12, 18, 34].

Definition 5.10 (Conditional pseudo Shannon entropy). A random variable Y_n jointly distributed with X_n is said to have (S,ε)-*conditional pseudo Shannon entropy* at least k, for any $S \in \mathbb{N}$ and $\varepsilon > 0$ if there exists a random variable Z_n jointly distributed with X_n such that

1. $H(Z_n|X_n) \geq k$, and
2. (X_n, Y_n) and (X_n, Z_n) are (S,ε)-indistinguishable.

Now suppose $Y = \{Y_n\}_{n\in\mathbb{N}}$ is an ensemble of random variables jointly distributed with another ensemble of distributions $X = \{X_n\}_{n\in\mathbb{N}}$. For any $S = S(n) > n$ and $\varepsilon = \varepsilon(n) > 0$, Y is said to have (S,ε)-*conditional pseudo Shannon entropy* at least $k = k(n)$ given X if there exists another ensemble of distributions Z jointly distributed with X such that for all sufficiently large n, Y_n has $(O(S(n)),\varepsilon(n))$-conditional pseudo Shannon entropy at least $k(n)$.

The following is the variant of pseudoentropy that we are looking for in this subsection and was introduced by Haitner *et al.* [12].

Definition 5.11 (Next-bit pseudo Shannon entropy). An ensemble of random variables $X = \{X_n\}_{n\in\mathbb{N}}$, where $X_n = (X_n^1, X_n^2, \cdots, X_n^n)$ takes values in Σ^n, has (S,ε)-*next-bit pseudo Shannon entropy* for any $S = S(n) > n$ and $\varepsilon = \varepsilon(n) > 0$ at least $k = k(n)$, denoted as $H^{next,S,\varepsilon}(X) \geq k$ if there exists an ensemble of random variables $Y = \{Y_n\}_{n\in\mathbb{N}}$ where $Y_n = (Y_n^1, Y_n^2, \cdots, Y_n^n)$ takes values in Σ^n and for all sufficiently large n,

1. $\sum_i H(Y_n^i|X_n^1, \cdots, X_n^{i-1}) \geq k(n)$, and
2. for all $1 \leq i \leq n$, $(X_n^1, \cdots, X_n^{i-1}, X_n^i)$ and $(X_n^1, \cdots, X_n^{i-1}, Y_n^i)$ are $(O(S(n)),\varepsilon(n))$-indistinguishable.

Later, Vadhan and Zheng [34] provided an alternative characterization of conditional pseudo Shannon entropy by showing an equivalence between it and *KL-hardness* (defined below). We use this alternative characterization extensively for our purpose.

Definition 5.12 (KL-hardness). Suppose (X_n, Y_n) is a $\Sigma^n \times \Sigma$-valued random variable and π be any predictor. Then π is said to be a δ-*KL-predictor* of Y_n given X_n if $\mathbf{KL}(X_n, Y_n \| X_n, C_\pi) \leq \delta$ where $C_\pi(y|x) = \pi(x,y)$ for all $x \in \Sigma^n$ and $y \in \Sigma$.

Moreover, for any $S = S(n) > n$ and $\delta = \delta(n) > 0$, an ensemble $Y = \{Y_n\}_{n\in\mathbb{N}}$, where each Y_n is a random variable taking value in Σ, is said to be (S,δ)-*KL-hard* given another ensemble of random variables $X = \{X_n\}_{n\in\mathbb{N}}$, where each X_n takes value in Σ^n, if for all large enough n there is no predictor π of size at most $O(S(n))$ that is a $\delta(n)$-KL-predictor of Y_n given X_n.

The following theorem provides the equivalence among KL-hardness and conditional pseudo Shannon entropy of a distribution.

Theorem 5.13 ([34]). *For an ensemble $(X,Y) = \{(X_n,Y_n)\}_{n\in\mathbb{N}}$ where each (X_n,Y_n) is a $\Sigma^n \times \Sigma$-valued random variable and for any $\delta = \delta(n) > 0$, $\varepsilon = \varepsilon(n) > 0$,*

1. *If Y is (S,δ)-KL-hard given X, then for all large enough n, Y_n has $(S'(n),\varepsilon(n))$-conditional pseudo Shannon entropy at least $H(Y_n|X_n) + \delta(n) - \varepsilon(n)$, where $S'(n) = S(n)^{\Omega(1)}/poly(n,1/\varepsilon(n))$.*
2. *Conversely, if for all large enough n, Y_n has $(S(n),\varepsilon(n))$-conditional pseudo Shannon entropy at least $H(Y_n|X_n) + \delta(n)$, then for every $\sigma = \sigma(n) > 0$, Y is (S',δ')-KL-hard given X, where $S'(n) = \min\{S(n)^{\Omega(1)}/poly(n,\log(1/\sigma(n))),\Omega(\sigma(n)/\varepsilon(n))\}$ and $\delta'(n) = \delta(n) - \sigma(n)$.*

Now we are ready to state the main theorem of this subsection which conveys the fact that the distributions with high dimensions also have high next-bit pseudo Shannon entropy.

Theorem 5.14. *For an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where D_n is a distribution over Σ^n, for any $S = S(n) > n$ and $\varepsilon = \varepsilon(n) > 0$ such that $\varepsilon(n) \to 0$ as $n \to \infty$, if $\dim_{S,\varepsilon}(D) > 2s$, then $H^{next,S',\varepsilon}(D) > k$, where $k = k(n) = sn$ and $S'(n) = S(n)^{\Omega(1)}/poly(n,1/\varepsilon(n))$.*[7]

Proof. For the sake of contradiction, let us assume that $H^{next,S',\varepsilon}(D) \le k$. Thus there exists a subset $S \subseteq \mathbb{N}$ of cardinality equals to the cardinality of $\mathbb{N}$ such that for all $n \in S$, Item 1 and Item 2 of Definition 5.11 do not hold simultaneously. On the other hand $\dim_{S,\varepsilon}(D) > 2s$ and thus by following the proof of Theorem 3.4, we can say that there exists a $c \in (0,1)$ such that $\text{unpred}_{S^c,\varepsilon}(D) = t > 2s$. Now consider some large enough n belongs to S and break D_n into 1-bit blocks, i.e., $D_n = (D_n^1, D_n^2, \cdots, D_n^n)$.

For any predictor $\pi : \Sigma^* \times \Sigma \to [0,1]$, let us define $\pi_i : \Sigma^{i-1} \times \Sigma \to [0,1]$ such that $\pi(x,a) = \pi_i(x,a), \forall_{x\in\Sigma^{i-1},a\in\Sigma}$ for $1 \le i \le n$. Then,

$$
\begin{aligned}
&\sum_{i=1}^{n} \mathbf{KL}((D_n^1,\cdots,D_n^{i-1}),D_n^i \| (D_n^1,\cdots,D_n^{i-1}),\pi_i) \\
&= \sum_{i=1}^{n} \Big[\sum_{w_i\in\Sigma^i} -\log(\pi_i(w_i[1\cdots i-1],w_i[i]))Pr[w_i] - H(D_n^i|D_n^1\cdots D_n^{i-1}) \Big] \\
&= \sum_{i=1}^{n} \Big[\sum_{w_i\in\Sigma^i} loss(\pi(w_i[1\cdots i-1],w_i[i]))Pr[w_i] - H(D_n^i|D_n^1\cdots D_n^{i-1}) \Big] \\
&= \sum_{w\in\Sigma^n} \text{Loss}(\pi,w)D_n(w) - H(D_n)
\end{aligned}
$$

where the last equality follows from the chain rule of Shannon entropy (Proposition 5.2).

Now the definitions of next-bit pseudo Shannon entropy, conditional pseudo Shannon entropy and KL-predictor along with Item 1 of Theorem 5.13 imply that for any $n \in S$, there exists a predictor π of size at most $O((S(n))^c)$ such that

$$\sum_{w\in\Sigma^n} \text{Loss}(\pi,w)D_n(w) \le (s+\varepsilon(n))n.$$

[7] Note that $\Omega(1)$ constant in the expression of $S'(n)$ here is related to that appeared in Item 1 of Theorem 5.13, but not exactly the same.

Hence for any $\varepsilon_1 > 0$,

$$\begin{aligned} D_n[\text{LossRate}(\pi, w) \geq (t-\varepsilon_1)] &= D_n[\text{Loss}(\pi, w) \geq (t-\varepsilon_1)n] \\ &\leq \frac{\sum_{w\in\Sigma^n} \text{Loss}(\pi, w) D_n(w)}{(t-\varepsilon_1)n} \quad \text{by Markov inequality} \\ &\leq \frac{s+\varepsilon(n)}{t-\varepsilon_1}. \end{aligned}$$

Thus,

$$D_n[\text{LossRate}(\pi, w) \leq (t-\varepsilon_1)] \geq 1 - \frac{s+\varepsilon(n)}{t-\varepsilon_1}.$$

As $\text{unpred}_{S^c,\varepsilon}(D) = t$, by the definition, it implies that for all but finitely many n belong to the set S,

$$1 - \frac{s+\varepsilon(n)}{t-\varepsilon_1} \leq \frac{1}{2} + \varepsilon(n).$$

For all large enough n, the above inequality gives us $t \leq 2s + \varepsilon_2$, for any $\varepsilon_2 > \varepsilon_1$ because $\varepsilon(n) \to 0$ as $n \to \infty$. Hence $\text{unpred}_{S^c,\varepsilon}(D) \leq 2s$ which is a contradiction and this completes the proof. □

The following weak converse can easily be proven.

Theorem 5.15. *For an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where D_n is a distribution over Σ^n, for any $S = S(n) > n$ and $\varepsilon = \varepsilon(n) > 0$, if $H^{next,S,\varepsilon}(D) > sn$, then for any $\varepsilon' > 0$ $\dim_{S',\varepsilon}(D) > s - \frac{1}{2} - \varepsilon'$, where $S'(n) = \min\{(S(n))^{\Omega(1)}/poly(n), 1/\varepsilon(n)^{\Omega(1)}\}$.*[8]

Proof. Suppose an ensemble of distributions D is given such that $H^{next,S,\varepsilon}(D) > sn$. Now take any large enough n and break D_n into 1-bit blocks, i.e., $D_n = (D_n^1, D_n^2, \cdots, D_n^n)$. Let us denote $\dim_{S',\varepsilon}(D) = t$ and thus by Theorem 3.4, there exists a constant $c > 1$ such that $\text{unpred}_{S'^c,\varepsilon}(D) \leq t$. This implies that for infinitely many n, there exists a predictor π of size at most $O(S'^c(n))$ such that for any $\varepsilon_1 > 0$,

$$D_n[\text{LossRate}(\pi, w) \leq (t+\varepsilon_1)] > \frac{1}{2} + \varepsilon(n).$$

Thus we can write the following,

$$\sum_{w\in\Sigma^n} \text{Loss}(\pi, w) D_n(w) \leq (t+\varepsilon_1)n + (\frac{1}{2} - \varepsilon(n))n.$$

We have already seen that for any predictor $\pi : \Sigma^* \times \Sigma \to [0,1]$,

$$\sum_{i=1}^{n} \mathbf{KL}((D_n^1, \cdots, D_n^{i-1}), D_n^i \| (D_n^1, \cdots, D_n^{i-1}), \pi_i) = \sum_{w\in\Sigma^n} \text{Loss}(\pi, w) D_n(w) - H(D_n).$$

Now the definitions of next-bit pseudo Shannon entropy, conditional pseudo Shannon entropy and KL-predictor along with Item 2 of Theorem 5.13 imply that for all large enough n, for every $\sigma(n) > 0$,

$$\sum_{w\in\Sigma^n} \text{Loss}(\pi, w) D_n(w) > (s - \sigma(n))n.$$

[8] Note that $\Omega(1)$ constant of the term $S(n)^{\Omega(1)}$ here is related to that appeared in Item 2 of Theorem 5.13, but not exactly the same.

Hence,

$$(s-\sigma(n))n < (t+\varepsilon_1)n + (\frac{1}{2}-\varepsilon(n))n$$
$$\Rightarrow s < t + \frac{1}{2} + \varepsilon_2$$

for any $\varepsilon_2 > \varepsilon_1$, for infinitely many large enough n and now setting $S'(n)$ to be such that $S'(n)^c$ equals to the value of $S'(n)$ mentioned in Item 2 of Theorem 5.13 concludes the proof. □

It is natural to allow $S(n)$ to be $n^{O(1)}$ and $\varepsilon(n)$ to be $n^{-O(1)}$ and then consider the definitions of conditional pseudo Shannon entropy, next-bit pseudo Shannon entropy (denoted as $H^{next}(D)$) and KL-hardness. For details we refer the reader to [34]. Now we use the equivalence between conditional pseudo Shannon entropy and KL-hardness from [34] to derive the following corollary of Theorem 5.14.

Corollary 5.16. *For any ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where D_n is a distribution over Σ^n, and for any $s \in [0,1]$, if $\dim(D_n) > 2s$, then $H^{next}(D) > k$ for $k = k(n) = sn$.*

In a similar fashion, we get the following as a corollary of Theorem 5.15.

Corollary 5.17. *For any $\varepsilon > 0$, $s \in [0,1]$ and $k = k(n) = sn$, for any ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where D_n is a distribution over Σ^n, if $H^{next}(D) > k$, then $\dim(D) > s - \frac{1}{2} - \varepsilon$.*

6. Pseudorandom Extractors & Lower Bound

We now introduce the notion of *pseudorandom extractor* similar to the notion of randomness extractor. Intuitively, a *randomness extractor* is a function that outputs almost random (statistically close to uniform) bits from weakly random sources, which need not be close to the uniformly random source. Two distributions X and Y on a set Λ are said to be $\varepsilon-$*close* (statistically close) if $\max_{S\subseteq\Lambda}\{|Pr[X\in S]-Pr[Y\in S]|\} \leq \varepsilon$ or equivalently $\frac{1}{2}\sum_{x\in\Lambda}|Pr[X=x]-Pr[Y=x]| \leq \varepsilon$.

Definition 6.1 (Deterministic Randomness Extractor). For any $\varepsilon = \varepsilon(n) > 0$, a family of functions $E = \{E_n\}_{n\in\mathbb{N}}$ where $E_n : \Sigma^n \to \Sigma^{m(n)}$ is said to be a deterministic ε-extractor for a class of ensemble of distributions $\mathscr{C}$ if for every ensemble of distributions $X = \{X_n\}_{n\in\mathbb{N}}$ in $\mathscr{C}$, where X_n is on Σ^n, for all large enough n, the distribution $E_n(X_n)$ is $\varepsilon(n)$-close to $U_{m(n)}$.

Likewise, a seeded ε-extractor is defined and the only difference is that now it takes a $d(n)$-bit string chosen according to $U_{d(n)}$, as an extra input. Before going further, we mention that for ease of presentation, now onwards we will only talk about the definitions and results derived so far related to pseudorandomness and dimension by considering all polynomial size circuits and inverse polynomial bias. We now define the notion of a pseudorandom extractor, the purpose of which is to extract out pseudorandom distribution from a given distribution.

Definition 6.2 (Pseudorandom Extractor). A family of functions $E = \{E_n\}_{n\in\mathbb{N}}$ where $E_n : \Sigma^n \to \Sigma^{m(n)}$ is said to be a *deterministic pseudorandom extractor* for a class of ensemble of distributions $\mathscr{C}$ if for every ensemble of distributions $X = \{X_n\}_{n\in\mathbb{N}}$ in $\mathscr{C}$, where X_n is on Σ^n, $E(X)$ is pseudorandom.

A family of functions $E = \{E_n\}_{n\in\mathbb{N}}$ where $E_n : \Sigma^n \times \Sigma^{d(n)} \to \Sigma^{m(n)}$ is said to be a *seeded pseudorandom extractor* for a class of ensemble of distributions $\mathscr{C}$ if for every ensemble of distributions $X = \{X_n\}_{n\in\mathbb{N}}$ in $\mathscr{C}$, $E(X, U_d)$ is pseudorandom.

In this section, we will concentrate on the class of distributions having dimension at least s. It is clear from the results stated in Section 5.2 that this class of distribution is a strict superset of the class of distributions with HILL-type pseudo min-entropy at least sn, for which any randomness extractor will act as a pseudorandom extractor [4]. Thus it is natural to ask the following.

Question 6.3. For any $s \in (0,1]$, does there exist a deterministic/seeded pseudorandom extractor for the class of ensemble of distributions having dimension at least s?

Just like the case of randomness extraction, one can easily argue that deterministic pseudorandom extraction is not possible[9]. The next natural question is what the lower bound on the seed length will be. We answer this question in the following theorem.

Theorem 6.4. *Suppose for any $s \in (0,1]$, $E = \{E_n\}_{n\in\mathbb{N}}$ where $E_n : \Sigma^n \times \Sigma^{d(n)} \to \Sigma^{m(n)}$ be a seeded pseudorandom extractor for the class of ensemble of distributions having dimension at least s and for some $\delta > 0$, $m(n) = (sn)^\delta$. Then $d(n) = \Omega(\log n)$.*

Proof. For the sake of contradiction, let us assume that $d(n) = o(\log n)$. Now by doing a walk according to the output distribution on an odd-length cycle, we achieve the following claim.

Claim. There is a deterministic $\frac{1}{\sqrt[4]{m}}$-extractor $E' = \{E'_m\}_{m\in\mathbb{N}}$ where $E'_m : \Sigma^m \to \Sigma^{\frac{\log m}{4}}$ for all pseudorandom ensemble of distributions.

This claim follows from the stronger result stated in Theorem 6.8. Now construct the following function $Ext_n : \Sigma^n \times \Sigma^{d(n)} \to \Sigma^{c\log n}$ for some constant $c > 0$ such that $Ext_n(x,y) = E'_n(E_n(x,y))$ for all $x \in \Sigma^n, y \in \Sigma^{d(n)}$. The function Ext is a seeded $\frac{1}{(sn)^{\delta/4}}$-extractor with $d(n) = o(\log n)$, but it is well known due to [30](Theorem 1.9) that any such randomness extractor must satisfy $d(n) = \Omega(\log n)$ and hence we get a contradiction. □

However, the question on constructing an *explicit* or polynomial time computable seeded pseudorandom extractor with seed length $O(\log n)$ is still open and next, we formally pose this question.

Question 6.5. For any $s \in (0,1]$, can one construct a seeded pseudorandom extractor $E = \{E_n\}_{n\in\mathbb{N}}$ where $E_n : \Sigma^n \times \Sigma^{d(n)} \to \Sigma^{m(n)}$ in polynomial time, for the class of ensemble of distributions having dimension at least s such that $m(n) = (sn)^\delta$ for some $\delta > 0$ and $d(n) = O(\log n)$?

Note that it is important to consider dimension in the statement of Question 6.3 and 6.5, because if we consider strong dimension instead of dimension then sometimes it might be just impossible to extract out pseudorandom distributions. For example one can consider the ensemble of distributions mentioned in the proof of Theorem 4.9 where however strong dimension is 1, as for infinitely many n, support of D_n contains just a single string, thus one cannot hope to get any pseudorandom distribution out of it. In the next part of this section, we will see a special type of nonpseudorandom source and give an explicit construction of deterministic pseudorandom extractor for that particular type of source. Before proceeding further, we want to mention that it is also very interesting to consider *nonpseudorandom distributions samplable by poly-size circuits* and we will discuss on the existence of extractor for that particular source in Section 6.2.

6.1. Deterministic Pseudorandom Extractor for Nonpseudorandom Bit-fixing Sources

In Section 4 while proving Theorem 4.3, we have introduced a special type of nonpseudorandom distribution which looks similar to the (n,k)-*bit-fixing source* defined as a distribution X over Σ^n such that there exists a subset $I = \{i_1,\ldots,i_k\} \subseteq \{1,\ldots,n\}$ where all the bits at the indices of I are independent and uniformly chosen and rest of the bits are completely fixed. This distribution was introduced by Chor *et al.*[6]. Now we define an analogous notion for the class of nonpseudorandom distributions, which we term *nonpseudorandom bit-fixing sources*.

Definition 6.6 (Nonpseudorandom Bit-fixing Source). Let $s \in (0,1)$. An ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$, where D_n is an distribution over Σ^n, with dimension s is an (n,s)-*nonpseudorandom bit-fixing source* if for all n there

[9]Suppose $E = \{E_n\}_{n\in\mathbb{N}}$ with $E_n : \Sigma^n \to \Sigma$ is a deterministic pseudorandom extractor, then for all n, there exists $x \in \Sigma$ such that $|E_n^{-1}(x)| \geq 2^{n-1}$. Thus E is not a pseudorandom extractor for a source D that is a uniform distribution on $E_n^{-1}(x)$ for all n and by Lemma 5.8, $\dim(D) \geq s$ for any $s < 1$.

exists a subset $I = \{i_1, \ldots, i_{\lceil sn \rceil}\} \subseteq \{1, \ldots, n\}$ such that all the bits at the indices of I come from a pseudorandom distribution and rest of the bits are fixed.

We devote the rest of the section to achieve an affirmative answer to the question of constructing deterministic pseudorandom extractor for the nonpseudorandom bit-fixing sources. For this purpose, we show that a careful analysis of the technique used in the construction of the deterministic randomness extractor for bit-fixing random sources by Gabizon, Raz and Shaltiel [9] will lead us to the desired deterministic pseudorandom extractor.

Theorem 6.7. *For any $s \in (0,1]$, there is an explicit deterministic pseudorandom extractor $E = \{E_n : \Sigma^n \to \Sigma^{m(n)}\}_{n \in \mathbb{N}}$, for all (n,s)-nonpseudorandom bit-fixing sources having polynomial-size support where $m(n) = (sn)^{\Omega(1)}$.*

We first extract $O(\log sn)$ amount of almost random bits and then use the same as seed in the seeded extractor. To use the seeded extractor, we modify the source such that it becomes independent of the random bits extracted. Before going into the exact details of the proof, we first discuss the ingredients required in the proof of the theorem.

6.1.1. Pseudorandom walk and extracting a few random bits

Kamp and Zuckerman [20] use a technique of random walk on odd-length cycles to extract almost random bits from a bit-fixing source. We adapt this to extract $O(\log sn)$ almost random bits from a (n,s)-nonpseudorandom bit-fixing source.

Theorem 6.8. *Let $s \in [0,1]$, $k = \lceil sn \rceil$. Then there is a deterministic $\frac{1}{\sqrt[4]{k}}$-extractor $E = \{E_n : \Sigma^n \to \Sigma^{\frac{\log k}{4}}\}_{n \in \mathbb{N}}$ for all (n,s)-nonpseudorandom bit-fixing sources.*

We prove the above theorem using the property of *pseudorandom walk* together with the fact that the second largest eigenvalue of a l length odd cycle is $\cos(\pi/l)$. Note that a corollary of the above theorem is the claim used in the proof of Theorem 6.4. Before proving the above theorem, we state two lemmas required for the proof. The first is a very special case of a lemma given in [20] which was restated in [9].

Lemma 6.9 ([9]). *Let $n \in \mathbb{N}$, $k \leq n$ and $s \in [0,1]$. Suppose G is an odd length cycle having M vertices and having second largest eigenvalue λ. If we take a walk on G according to the bits from a (n,k)-bit-fixing source, starting from any fixed vertex, then at the end of the n step of the walk, the distribution P on the vertices will be $\frac{1}{2}\lambda^k\sqrt{M}$-close to U_M.*

Now we prove a similar result for (n,s)-nonpseudorandom bit-fixing source using the property of *pseudorandom walk*. The idea of pseudorandom walk was also used previously in the domain of space bounded computation by Reingold *et al.* [31].

Lemma 6.10. *Let $s \in [0,1]$ and $k = \lceil sn \rceil$. Let G be an odd length cycle having M vertices and having second largest eigenvalue λ. If we take a walk on G according to the bits from a (n,s)-nonpseudorandom bit-fixing source starting from any fixed vertex, then for all large enough n, at the end of the n step of the walk, the distribution Q on the vertices will be $\frac{1}{2}(\lambda^k + \sqrt{M}\varepsilon(n))\sqrt{M}$-close to U_M, where M is polynomial in n and $\varepsilon(n) < 1/n^c$ for any constant $c > 0$.*

Proof. Let π be the stationary distribution on the vertices and since we consider an odd length cycle (a 2-regular graph), the stationary distribution is the uniform distribution on M vertices. Suppose we take a n step walk on the graph G starting from any vertex v according to the bits from a (n,k)-bit-fixing source, where $k = \lceil sn \rceil$ and the probability vector on the vertices at the end of the walk is $P = \begin{pmatrix} p_1 & p_2 & \ldots & p_M \end{pmatrix}$. Now we take a n step walk on the graph G starting from the same vertex v according to the bits from a (n,s)-nonpseudorandom bit-fixing source and the probability vector on the vertices at the end of the walk is $Q = \begin{pmatrix} q_1 & q_2 & \ldots & q_M \end{pmatrix}$, where $\forall_{1 \leq i \leq M}$, $q_i \leq p_i + \varepsilon(n)$ and $\varepsilon(n) < 1/n^c$ for any constant $c > 0$. This bound on q_i can be justified as follows.

Note that the only difference between (n,s)-nonpseudorandom bit-fixing source and (n,k)-bit-fixing source is that on the set I, in (n,k)-bit-fixing source, we have U_k instead of pseudorandom distribution (say D) on Σ^k. Also observe that actually P and Q are the distributions on vertices at the end of a k step walk, where the walk was started from the vertex v and done according to the bits coming from U_k and D respectively, because a step according to a fixed bit will not change the output distribution and in a (n,k)-bit-fixing source (also in a (n,s)-nonpseudorandom bit-fixing source), all the $n-k$ bits are fixed. For a step according to a fixed bit gives rise to a transition matrix that is actually a permutation matrix and thus it leaves the distance from uniform unchanged. Hence, if the bound on $q_i, \forall_{1\leq i\leq M}$ is not true then we can use this k step walk on G as a polynomial (polynomial in k) time algorithm to distinguish between U_k and D. Thus,

$$||q-\pi||^2 = \sum_{i=1}^{M}(q_i - \frac{1}{M})^2 \leq \sum_{i=1}^{M}(p_i+\varepsilon(n)-\frac{1}{M})^2 = ||p-\pi||^2 + M\varepsilon(n)^2 \leq \lambda^{2k} + M\varepsilon(n)^2 \leq (\lambda^k + \sqrt{M}\varepsilon(n))^2.$$

□

The above lemma together with the fact that the second largest eigenvalue of a l length odd cycle is $\cos(\pi/l)$, implies Theorem 6.8.

Proof of Theorem 6.8. Let us take an odd cycle G with $M = \sqrt[4]{k}$ vertices. The second largest eigenvalue of G is $\cos(\frac{\pi}{\sqrt[4]{k}})$. Now take a walk starting from a fixed vertex of G according to the bits from (n,s)-nonpseudorandom bit-fixing source and finally output the vertex number of the graph G. Thus we get $\frac{\log k}{4}$ bits. From Lemma 6.10, we reach distance $\frac{1}{2}((\cos(\frac{\pi}{\sqrt[4]{k}}))^k + \sqrt[8]{k}\varepsilon(n))\sqrt[8]{k}$ from uniform.

By the Taylor expansion of the cosine function, for $0 < x < 1$, $\cos(x) < 1 - \frac{x^2}{2} + \frac{x^4}{24}$.
Therefore, $\left(\cos\left(\frac{\pi}{\sqrt[4]{k}}\right)\right)^k < (1-\frac{\pi^2}{4\sqrt{k}})^k < (\exp^{-\frac{\pi^2}{4}})^{\sqrt{k}} < 4^{-\sqrt{k}}$. Hence, $\frac{1}{2}((\cos(\frac{\pi}{\sqrt[4]{k}}))^k + \sqrt[8]{k}\varepsilon(n))\sqrt[8]{k} < \frac{1}{\sqrt[4]{k}}$. Thus we get distribution of $\frac{\log k}{4}$ bit strings which is $\frac{1}{\sqrt[4]{k}}$-close to uniform in statistical distance. □

6.1.2. Sampling and Partitioning with a short seed

Here we restate some of the results on sampling and partitioning used in construction of deterministic extractor for bit-fixing sources from [9]. Let $S \subseteq [n]$ be some subset of size k. Now we consider a process of generating a subset $T \subseteq [n]$ such that $k_{min} \leq |S \cap T| \leq k_{max}$ and this process is known as *Sampling*.

Definition 6.11. A function $Samp : \Sigma^t \to P([n])$ is called a $(n,k,k_{min},k_{max},\delta)$-sampler if for any subset $S \subseteq [n]$, where $|S| = k$, $Pr_{w\in_R U_t}[k_{min} \leq |Samp(w) \cap S| \leq k_{max}] \geq 1-\delta$

Now consider a similar process known as *Partitioning*, the task of which is to partition $[n]$ into m distinct subsets $T_1, T_2, \cdots, T_m$ such that for every $1 \leq i \leq m$, $k_{min} \leq |S \cap T_i| \leq k_{max}$.
According to [9], the above two processes can be performed using only *a few* random bits.

Lemma 6.12 ([9]). *For any constant $0 < \alpha < 1$, there exist constants $c > 0, 0 < b < 1$ and $\frac{1}{2} < e < 1$ such that for any $n \geq 16$ and $k \geq (\log n)^c$, there is an explicit construction of a function $Samp : \Sigma^t \to P([n])$ which is a $(n,k,\frac{k^e}{2},3k^e,O(k^{-b}))$-sampler, where $t = \alpha \log k$.*

Lemma 6.13 ([9]). *For any constant $0 < \alpha < 1$, there exist constants $c > 0, 0 < b < 1$ and $\frac{1}{2} < e < 1$ such that for any $n \geq 16$ and $k \geq (\log n)^c$, there is an explicit construction that uses only $\alpha \log k$ random bits and partition $[n]$ into $m = O(k^b)$ many subsets $T_1, T_2, \cdots, T_m$ such that for any subset $S \subseteq [n]$, where $|S| = k$, $Pr[\forall 1 \leq i \leq m, \frac{k^e}{2} \leq |T_i \cap S| \leq 3k^e] \geq 1 - O(k^{-b})$.*

6.1.3. Generating an independent seed

In this subsection, we see the way of obtaining a short seed from a nonpseudorandom bit-fixing source so that we can use them in a seeded pseudorandom extractor to extract out almost all the pseudorandom part from the source. The main problem of using this short seed in a seeded pseudorandom extractor is that the already obtained seed is dependent on the main distribution. Now we describe that this problem can be removed in the case of nonpseudorandom bit-fixing sources. Even though the result is analogous to [9], the proofs differ in essential details.

Definition 6.14 (Seed Obtainer). A family of functions $F = \{F_n : \Sigma^n \to \Sigma^n \times \Sigma^{d(n)}\}_{n\in\mathbb{N}}$ is said to be a (s,s',ρ)-seed obtainer ($s' \leq s$) if for every (n,s)-nonpseudorandom bit-fixing source $X = \{X_n\}_{n\in\mathbb{N}}$, the distribution $R = \{R_n = F_n(X_n)\}_{n\in\mathbb{N}}$ can be written as $R = \eta Q + \sum_a \alpha_a R_a$[10] for $\eta = \eta(n) > 0$, $\alpha_a = \alpha_a(n) > 0$ and $\eta(n) + \sum_a \alpha_a(n) = 1$ such that $\eta(n) \leq \rho(n)$ and for every a, there exists a (n,s')-nonpseudorandom bit-fixing source Z_a such that for all large enough n, $\{R_a\}_n$ is $\rho(n)$-close to $\{Z_a\}_n \otimes U_{d(n)}$.

In the above definition, by $\{Z_a\}_n \otimes U_{d(n)}$, we mean the product of two distributions $\{Z_a\}_n$ and $U_{d(n)}$. From the above definition it is clear that given a seed obtainer and a seeded pseudorandom extractor for nonpseudorandom bit-fixing sources, we can easily construct a deterministic pseudorandom extractor. The following theorem provides us the details of such construction, where the correctness follows from the properties of our proposed notion of dimension described in Section 4.

Theorem 6.15. *Suppose $F = \{F_n : \Sigma^n \to \Sigma^n \times \Sigma^{d(n)}\}_{n\in\mathbb{N}}$ is a (s,s',ρ)-seed obtainer, where $\rho(n) \leq \frac{1}{(sn)^c}$ for some constant $c > 0$ and $E' = \{E'_n : \Sigma^n \times \Sigma^{d(n)} \to \Sigma^{m(n)}\}_{n\in\mathbb{N}}$ is a seeded pseudorandom extractor for (n,s')-nonpseudorandom bit-fixing sources, where $m(n) = (sn)^{\Omega(1)}$. Then the function $E = \{E_n : \Sigma^n \to \Sigma^{m(n)}\}_{n\in\mathbb{N}}$ defined as $E_n(x) = E'_n(F_n(x))$ for all $x \in \Sigma^n$, is a deterministic pseudorandom extractor for (n,s)-nonpseudorandom bit-fixing sources.*

Proof. By the definition of the seed obtainer, we can write $E(X) = \eta E'(Q) + \sum_a \alpha_a E'(R_a) = \eta E'(Q) + (1-\eta)Y$, for some ensemble of distributions Y. Now by Lemma 4.5, for all a, $E'(R_a)$ is pseudorandom and as a consequence, by Lemma 4.6, Y is pseudorandom as well. Then using Lemma 4.5, we can argue that $E(X)$ is also pseudorandom as $\eta \leq \frac{1}{(sn)^c}$, for some constant $c > 0$. □

Now we give an explicit construction of (s,s',ρ)-seed obtainer, which is crucial in the later part of this paper. To understand the notion of *sampler* used in the following theorem, the readers may refer to the last subsection.

Theorem 6.16. *Let $Samp = \{Samp_n : \Sigma^{t(n)} \to P([n])\}_{n\in\mathbb{N}}$ where $Samp_n$ be a $(n, \lceil sn \rceil, \lceil s_1 n \rceil, \lceil s_2 n \rceil, \delta(n))$-sampler and $E = \{E_n : \Sigma^n \to \Sigma^{m(n)}\}_{n\in\mathbb{N}}$ with $m(n) > t(n)$ be a deterministic ε-extractor for (n,s_1)-nonpseudorandom bit-fixing sources, where $\varepsilon(n) < 1/n^c$ for any constant $c > 0$. Then there is an explicit (s,s',ρ)-seed obtainer $F = \{F_n : \Sigma^n \to \Sigma^n \times \Sigma^{d(n)}\}_{n\in\mathbb{N}}$, where $d(n) = m(n) - t(n)$, $s' = s - s_2$, and $\rho(n) = \max\{\varepsilon(n) + \delta(n), \sqrt{\varepsilon(n)}2^{t(n)+1}\}$.*

The construction of the seed obtainer is the same as that mentioned in [9], however the proof requires a slightly different argument.

Proof. The construction of F mentioned in the theorem is as follows:

1. Given $x \in \Sigma^n$, compute $E_n(x)$. Denote the first $t(n)$ bits of $E_n(x)$ by $E^1_n(x)$ and the last $(m(n) - t(n))$ bits by $E^2_n(x)$;
2. Compute $Samp_n(E^1_n(X))$ and denote it as T;
3. Let $x' = x_{[n]\setminus T}$ and $y = E^2_n(x)$. If $|x'| < n$, pad it with zeros to get n-bit long string. Now output x', y.

[10] It means for all n, $R_n = \eta(n)Q_n + \sum_a \alpha_a(n)\{R_a\}_n$

Note that the above construction is the same as the construction of seed obtainer given in [9]. However, the proof is not the same and more specifically the proof of the next claim differs from that of the similar claim made in [9]. Here, in the proof we use the properties of pseudorandomness and the fact that the distribution under consideration has polynomial-size support.

Let X be a (n,s)-nonpseudorandom bit-fixing source and now consider any large enough n and let I be the set of indices at which the bits are not fixed. For a string $a \in \Sigma^{t(n)}$, T_a denotes $Samp_n(a)$ and T'_a denotes $[n] \setminus Samp_n(a)$. Given a string $x \in \Sigma^n$, x_a denotes x_{T_a} and x'_a denotes n-bit string obtained by padding $x_{T'_a}$ with zeros. Let $X' = X'_{E^1_n(X_n)}$ and $Y = E^2_n(X_n)$. We say that a string $a \in \Sigma^{t(n)}$ *correctly splits* X_n if $\lceil s_1 n \rceil \leq |I \cap T_a| \leq \lceil s_2 n \rceil$.

Claim. For every $a \in \Sigma^{t(n)}$ which correctly splits X_n, $(X'_a, E_n(X_n))$ is $\varepsilon(n)$-close to $(X'_a \otimes U_{m(n)})$, where $\varepsilon(n) < 1/n^c$ for any constant $c > 0$.

Proof. Let $|Samp_n(a)| = l$. Given a string $\sigma \in \Sigma^l$ and a string $\sigma' \in \Sigma^{n-l}$, we define $[\sigma;\sigma']$ as follows:
Suppose l indices of T_a are $i_1 < \cdots < i_l$ and the $(n-l)$ indices of T'_a are $i'_1 < \cdots < i'_{n-l}$. The string $[\sigma;\sigma'] \in \Sigma^n$ is defined as:

$$[\sigma;\sigma']_i = \begin{cases} \sigma_j & i \in T_a \text{ and } i_j = i \\ \sigma'_j & i \in T'_a \text{ and } i'_j = i \end{cases}$$

In this notation, we denote $X_n = [X_a; X'_a]$. Now consider the distribution $(X'_a, E_n(X_n)) = (X'_a, E_n([X_a; X'_a]))$. For every $b \in \Sigma^{n-l}$, we consider the event $\{X'_a = b\}$. As a correctly splits X_n, there are at least $\lceil s_1 n \rceil$ "good" indices in T_a. Now fix some $b \in \Sigma^{n-l}$ such that $X_n[X'_a = b] > 0$.

Now we claim that for all subsets $B \subseteq \Sigma^{n-l}$ where $\forall_{b \in B}, X_n[X'_a = b] > 0$, there exists a $b' \in B$ such that the distribution $([X_a; X'_a] | X'_a = b')$ forms an (n, s_1)-nonpseudorandom bit-fixing source if for some constant $c > 0$, $\varepsilon'(n) \geq 1/n^c$ and

$$\sum_{b \in B} X_n[X'_a = b] > \varepsilon'(n).$$

For the sake of contradiction, let us assume that the above claim is not true. It means that there exists a subset $J \subseteq \Sigma^{n-l}$, where

i. $\forall_{b \in J}, X_n[X'_a = b] > 0$,
ii. $\sum_{b \in J} X_n[X'_a = b] > \varepsilon'(n)$ where $\varepsilon'(n) \geq 1/n^c$, for some constant $c > 0$, and also
iii. for all $b \in J$, the distributions $([X_a; X'_a] | X'_a = b)$ are not forming (n, s_1)-nonpseudorandom bit-fixing sources.

Now let us consider only the "good" positions which are $\lceil sn \rceil$ many in X and at least $\lceil s_1 n \rceil$ many in $([X_a; X'_a] | X'_a = b)$. So the above assumption implies that the ensemble of distributions formed by considering those $\lceil s_1 n \rceil$ bits (this part of the string b is denoted as $b_{\lceil s_1 n \rceil}$) in $([X_a; X'_a] | X'_a = b)$ is not pseudorandom, i.e., it has its corresponding distinguishing circuits C_b. If this is the case, then the circuit C (by hard-wiring the good random bits) corresponding to the following algorithm A, will act as a distinguishing circuit for the pseudorandom distribution P on $\lceil sn \rceil$ many bits; which is a contradiction. The algorithm A is as follows: on input $y \in \{0,1\}^{\lceil s_1 n \rceil}$, if $y_{\lceil s_1 n \rceil} = b_{\lceil s_1 n \rceil}$ for any $b \in J$, then return $C_b(y_{\lceil s_1 n \rceil})$; otherwise return 0 or 1 uniformly. And thus clearly,

$$|P[A[y] = 1] - U_{\lceil s_1 n \rceil}[A[y] = 1]| > 1/n^c.$$

Circuit C is nothing but the combination of all the circuits C_b, for $b \in J$, each of which is of polynomial size. Now as $\forall_{b \in J}, X_n[X'_a = b] > 0$ and by our assumption that the distribution under consideration has polynomial-size support (see statement of Theorem 6.7), the support of the subset J is at most polynomial. Hence the circuit C is of polynomial size. Note that this is the only place where we use the fact that the distribution under consideration is of polynomial-size support.

So, we can write,

$$\frac{1}{2}\sum_{b,c}|Pr[(X'_a,E(X))=(b,c)]-Pr_{(X'_a\otimes U_{m(n)})}[b,c]|$$
$$=\frac{1}{2}\sum_{b,c}|Pr[X'_a=b]Pr[E_n(X_n)=c|X'_a=b]-Pr[X'_a=b]Pr_{U_{m(n)}}[c]|\leq\varepsilon(n)$$

where $\varepsilon(n)<1/n^c$ for any constant $c>0$. The first inequality follows from the fact that we can split the sum in two parts one in which $([X_a;X'_a]|X'_a=b)$'s are not (n,s_1)-nonpseudorandom bit-fixing sources and another in which $([X_a;X'_a]|X'_a=b)$'s are at least (n,s_1)-nonpseudorandom bit-fixing sources. □

Next we mention a claim from [9] that makes comment on independence of the pair $(X'_a,E^2_n(X_n))$ conditioned on the event $E^1_n(X_n)=a)$.

Claim ([9]). For every fixed $a\in\Sigma^{t(n)}$ that correctly splits X_n, the distribution $((X'_a,E^2_n(X))\mid E^1_n(X)=a)$ is $\varepsilon(n)2^{t+1}$-close to $(X'_a\otimes U_{m(n)-t(n)})$.

Note that as a correctly splits X_n, so X'_a forms a $(n,s-s_2)$-nonpseudorandom bit-fixing source.

The rest of the proof follows directly from the proof of correctness of the construction of seed obtainer given in [9] with the following parameters $k=\lceil sn\rceil$, $k_{min}=\lceil s_1n\rceil$, $k_{max}=\lceil s_2n\rceil$. □

6.1.4. A seeded pseudorandom extractor

In this subsection, we discuss about how we can extract $(sn)^{\Omega(1)}$ many pseudorandom bits using $O(\log sn)$ random bits. In the next subsection, we will use this seeded pseudorandom extractor and the techniques discussed in the previous subsections, to construct deterministic extractor. The construction of seeded pseudorandom generator given in the proof of the following theorem is same as that of the seeded randomness extractor given in [9]. However, the analysis is quite different and uses some of the properties of dimension.

Theorem 6.17. *For an $s\in(0,1)$ and any constant $0<\alpha<1$, there exist constants $c>0,0<b<1$ such that there is an explicit function $E=\{E_n:\Sigma^n\times\Sigma^{d(n)}\to\Sigma^{m(n)}\}_{n\in\mathbb{N}}$ which acts as a seeded pseudorandom extractor for (n,s)-nonpseudorandom bit-fixing sources with $d(n)=\alpha\log sn$ and $m(n)=\Omega((sn)^b)$.*

Proof. Let X be a (n,s)-nonpseudorandom bit-fixing source and for some large enough n, x be a string sampled by X_n. The description of the extractor $E_n(x,y)$ is as follows:

1. According to Lemma 6.13 provided in Section 6.1.2, using y as seed, we obtain a partition of $[n]$ into $m(n)=\Omega((sn)^b)$ many sets $T_1,T_2,\cdots,T_{m(n)}$ with the parameter α;
2. For $1\leq i\leq m(n)$, compute $z[i]=\oplus_{j\in T_i}x[j]$;
3. Output $z=z[1]z[2]\cdots z[m(n)]$.

Let $I\subseteq[n]$ be the set of indices at which the bits are not fixed and let Z_n be the distribution of the output strings. We need to show that $Z=\{Z_n\}_{n\in\mathbb{N}}$ is pseudorandom.

Let A_n be the event $\{\forall i,|T_i\cap I|\neq 0\}$ and $A^c_n=\{\exists i,|T_i\cap I|=0\}$ be the complement event. According to Lemma 6.13, $Pr[A_n]\geq 1-O((sn)^{-b})$. Now we can write the output distribution as

$$Z_n=Pr[A_n](Z_n|A_n)+Pr[A^c_n](Z_n|A^c_n)$$

and hence due to Lemma 4.5, Z is pseudorandom. □

6.1.5. Deterministic pseudorandom extractor

Now it only remains to combine all the components we discussed so far to build the final deterministic pseudorandom extractor mentioned in Theorem 6.7. We first extract $O(\log sn)$ amount of almost random bits by Theorem 6.8 and then use the same as seed in the seeded extractor described in Theorem 6.17. To use the seeded extractor it is required to modify the source such that it becomes independent of the random bits extracted using Theorem 6.8. For that purpose, we use the technique developed in Section 6.1.3 and this concludes the proof of Theorem 6.7.

Proof of Theorem 6.7. Due to Lemma 6.12, we have a $(n, sn, \frac{(sn)^e}{2}, 3(sn)^e, (sn)^{-\Omega(1)})$-sampler $Samp_n : \Sigma^{t(n)} \to P([n])$, where $t(n) = \frac{\log sn}{32}$ and $e > \frac{1}{2}$. From Theorem 6.8, we have a deterministic $\frac{1}{\sqrt[4]{s'n}}$-extractor $E^* = \{E^*_n : \Sigma^n \to \Sigma^{m'}\}_{n\in\mathbb{N}}$ for (n, s')-nonpseudorandom bit-fixing sources where for all large enough n, $s'n \geq \frac{(sn)^e}{2}$ and $m'(n) = \frac{\log s'n}{4}$. Now we use Theorem 6.16 to get (s, s'', ρ)-seed obtainer $F = \{F_n : \Sigma^n \to \Sigma^n \times \Sigma^{m'(n)-t(n)}\}_{n\in\mathbb{N}}$ where for all large enough n, $s''n \geq 3(sn)^e$ and $\rho(n) = \frac{1}{(sn)^p}$, for some constant p. According to Theorem 6.17, we have a seeded pseudorandom extractor $E' = \{E'_n : \Sigma^n \times \Sigma^{d(n)} \to \Sigma^{m(n)}\}_{n\in\mathbb{N}}$ with $d(n) = \frac{\log sn}{32}$ and $m(n) = (sn - s''n)^{\Omega(1)}$ for $(n, s - s'')$-nonpseudorandom bit-fixing sources. Since $m'(n) = \frac{\log s'n}{4} \geq \frac{\log sn}{16} = t(n) + d(n)$, we use F and E' in Theorem 6.15 to construct a deterministic pseudorandom extractor $E = \{E_n : \Sigma^n \to \Sigma^{m(n)}\}_{n\in\mathbb{N}}$. For a large enough n, $m(n) = (sn - s''n)^{\Omega(1)} = (sn)^{\Omega(1)}$ and this completes the proof. □

6.2. Discussion on Pseudorandom Extractor for Nonpseudorandom Samplable Distributions

Another interesting special kind of source is samplable distributions studied by Trevisan and Vadhan [33]. In a natural way, one can extend the definition of samplable distribution to nonpseudorandom distribution as follows: for any $s \in (0, 1]$, an ensemble of distributions $D = \{D_n\}_{n\in\mathbb{N}}$ is said to be s-nonpseudorandom samplable by circuit of size $S = S(n) > n$ if for all large enough n, there exists a circuit C of size at most $S(n)$ that samples from D_n and $\dim(D) = s$. Observe that the negative results for deterministic randomness extractor in case of samplable distributions will also applicable for deterministic pseudorandom extractor in case of s-nonpseudorandom samplable distribution. By Lemma 5.8, if $H_\infty(D_n) \geq n - 1$ for all large enough n, then $\dim(D) \geq s$ for any $s < 1$. Now by applying the argument in [33], we get the following.

Theorem 6.18. *Suppose $E = \{E_n : \Sigma^n \to \Sigma\}_{n\in\mathbb{N}}$ is a family of functions computable in time $T(n)$ such that E is a deterministic pseudorandom extractor for ensemble of distributions that are s-nonpseudorandom samplable by circuit of size $S(n)$ for any $s < 1$. Then there is a language in DTIME$(T(n))$ of circuit complexity at least $\Omega(S(n))$.*

The existence of deterministic pseudorandom extractors implies separations between deterministic complexity classes and non-uniform circuit classes that are not yet known. So one might have to consider some complexity theoretic assumptions like in [33] to construct deterministic pseudorandom extractor. However, we do not think construction with such strong assumption like in [33] will be interesting in this case as it is known that certain hardness assumption already leads to a construction of optimal pseudorandom generator (See Section 7). Nevertheless, it is natural to ask the question of constructing explicit extractor using $O(\log n)$ amount of extra randomness. We do not know any such result so far, but in the next section we will see that if some distribution is samplable using very few ($O(\log n)$) random bits, then it is possible to extract out all the pseudorandom bits using extra $O(\log n)$ random bits.

7. Approaching Towards P=BPP

We now show that if there is an exponential time computable algorithm $G = \{G_n\}_{n\in\mathbb{N}}$ with $G_n : \Sigma^{O(\log n)} \to \Sigma^n$ where the output distribution has dimension s ($s > 0$), then this will imply P=BPP. We refer to this algorithm G as *optimal nonpseudorandom generator*. The proof of this is similar to the proof of Theorem 7.4 [27]. We start with some basic definitions.

A pseudorandom generator against a class of circuits is a function which takes a random seed as input and outputs a sequence of bits which is a pseudorandom distribution.

Definition 7.1 (Pseudorandom Generators). A function G is said to be a $l(n)$-pseudorandom generator if

1. $G = \{G_n\}_{n\in\mathbb{N}}$ with $G_n : \Sigma^{l(n)} \to \Sigma^n$
2. G_n is computable in $2^{O(l(n))}$ time
3. For sufficiently large n, $G_n(U_{l(n)})$ is $(n^2, 1/n)$-pseudorandom.

Definition 7.2 (Optimal Pseudorandom Generators). A function G is said to be an optimal pseudorandom generator if it is an $O(\log n)$-pseudorandom generator.

Nisan and Wigderson [27] showed that there is a connection between pseudorandom generators and *hard functions* in EXP:

Definition 7.3 (Hard Function). A family of functions $f = \{f_n\}_{n\in\mathbb{N}}$ where $f_n : \Sigma^n \to \Sigma$ is (S,ε)-hard for any $S = S(n) > n$ and $\varepsilon = \varepsilon(n) > 0$, if for all large enough n, for all circuits C of size at most $O(S(n))$,

$$U_n[C(x) = f_n(x)] \leq \frac{1}{2} + \varepsilon(n).$$

The following theorem shows that under the assumption of existence of hard function in EXP, optimal pseudorandom generator exists [27].

Theorem 7.4 ([27]). *There exists an optimal pseudorandom generators if and only if there is a language L in* EXP *and $\exists\delta > 0$ such that L is $(2^{\delta n}, 1/2^{\delta n})$-hard.*

The proof of the above theorem is constructive and thus we can explicitly convert optimal pseudorandom generators to the hard function and conversely. However this is still a very strong requirement and later Impagliazzo and Wigderson weakened it.

Theorem 7.5 ([19]). *Suppose there is a language L in* EXP *and $\exists\delta > 0$ such that L on inputs of length n cannot be solved by circuits of size at most $2^{\delta n}$. Then there exists a language L' in* EXP *and $\exists\delta' > 0$ such that L' is $(2^{\delta' n}, 1/2^{\delta' n})$-hard and as a consequence optimal pseudorandom generator exists.*

Now let us state and prove the main result of this section.

Theorem 7.6. *Consider any $s \in (0,1]$ and $c > 0$. If there exists an algorithm $G = \{G_n\}_{n\in\mathbb{N}}$ where $G_n : \Sigma^{c\log n} \to \Sigma^n$ computable in $2^{O(\log n)}$ such that* $\dim(\{G_n(U_{c\log n})\}) \geq s$, *then* P=BPP.

Proof. Suppose $X := \{G_n(U_{c\log n})\}_{n\in\mathbb{N}}$. If $\dim(X) = s > 0$, then for all large enough n, there must be a subset of indices $S \subseteq \{1,2,\cdots,n\}$ such that $|S| = \log n$ and for any $i \in S$, loss incurred by any polynomial-size predictor at i-th bit position is non-zero or in other words, for any polynomial-size circuit family $C = \{C_n\}_{n\in\mathbb{N}}$, $X[C_i(x_1,\cdots,x_{i-1}) = x_i] < 1$. Actually one can show a much stronger claim that there exists such a subset S such that $|S| = c\cdot n$, for some constant $c < 1$. Otherwise $\dim(X) = 0$. To prove this, suppose $|S| = o(n)$ and thus there exists a predictor π such that for every $w \in \Sigma^n$,

$$\text{Loss}(\pi, w) = \sum_{i=1}^{n} loss(\pi(w[1\ldots i-1],\ w[i])) = \sum_{i\in S} loss(\pi(w[1\ldots i-1],\ w[i])).$$

Now as $loss(\pi(w[1\ldots i-1],\ w[i])) \leq 1$, so for all large enough n, for any $\varepsilon < \frac{1}{2}$, $\text{LossRate}_\varepsilon(\pi, X_n) = 0$. Hence $\text{unpred}(X) = 0$ and by Corollary 3.7, $\dim(X) = 0$ as well.

Suppose S contains first $\log n$ many such indices. Also assume that $S = \{i_1, i_2, \cdots, i_{\log n}\}$ and $i_1 < i_2 < \cdots < i_{\log n}$. Now we define two languages L_0 and L_1 as follows: for $j = 0, 1$,

$$L_j := \{y \in \Sigma^{\log n - 1} | \exists x \in \Sigma^n \text{ in the support of } G_n \text{ and } x_S = jy\}$$

where jy denotes the string generated by concatenating j and y. First of all, note that as $i_1 \in S$, none of L_0 and L_1 is an empty set. Now clearly either L_0 or L_1 is the language that satisfies all the conditions of Theorem 7.5 [19]. Otherwise, there exists a predictor circuit of size at most $2^{\delta \log n}$, for some $\delta > 0$, i.e., polynomial in n, by which we can predict $i_{\log n}$-th bit position or loss incurred by that predictor at $i_{\log n}$-th bit position will be zero implying $i_{\log n} \notin S$ which is a contradiction. Thus either L_0 or L_1 can be used to construct an *optimal pseudorandom generator* and that eventually implies P=BPP. □

Acknowledgements

The authors thank Somenath Biswas for helpful discussions and comments. The second author thanks Andrej Bogdanov for suggesting the study of the notion of pseudoentropy. The authors also thank the anonymous reviewers for their helpful comments and suggestions. The second and third author would like to acknowledge the support of Research-I Foundation and the third author would also like to acknowledge the support of the European Research Council under the European Unions Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 616787.

References

[1] Manindra Agrawal, Diptarka Chakraborty, Debarati Das, and Satyadev Nandakumar. Dimension, pseudorandomness and extraction of pseudorandomness. In *35th IARCS Annual Conference on Foundation of Software Technology and Theoretical Computer Science, FSTTCS 2015, December 16-18, 2015, Bangalore, India*, pages 221–235, 2015.

[2] K. B. Athreya, J. M. Hitchcock, J. H. Lutz, and E. Mayordomo. Effective strong dimension, algorithmic information, and computational complexity. *SIAM Journal on Computing*, 37:671–705, 2007.

[3] K. B. Athreya and S. N. Lahiri. *Measure Theory and Probability Theory*. Springer Verlag, 2006.

[4] Boaz Barak, Ronen Shaltiel, and Avi Wigderson. Computational analogues of entropy. In Sanjeev Arora, Klaus Jansen, Jos D. P. Rolim, and Amit Sahai, editors, *RANDOM-APPROX*, volume 2764 of *Lecture Notes in Computer Science*, pages 200–215. Springer, 2003.

[5] Manuel Blum and Silvio Micali. How to generate cryptographically strong sequences of pseudo-random bits. *SIAM J. Comput.*, 13(4):850–864, November 1984.

[6] Benny Chor, Oded Goldreich, Johan Håstad, Joel Freidmann, Steven Rudich, and Roman Smolensky. The bit extraction problem or t-resilient functions, 1985.

[7] T. Cover. Universal gambling schemes and the complexity measures of Kolmogorov and Chaitin. Technical Report 12, Stanford University Department of Statistics, October 1974.

[8] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[9] Ariel Gabizon, Ran Raz, and Ronen Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed. In *45th Symposium on Foundations of Computer Science (FOCS 2004), 17-19 October 2004, Rome, Italy, Proceedings*, pages 394–403, 2004.

[10] Oded Goldreich. *The Foundations of Cryptography - Volume 1, Basic Techniques*. Cambridge University Press, 2001.

[11] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 28(2):270–299, 1984.

[12] Iftach Haitner, Omer Reingold, and Salil P. Vadhan. Efficiency improvements in constructing pseudorandom generators from one-way functions. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 437–446, 2010.

[13] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.

[14] J. M. Hitchcock. Effective Fractal Dimension Bibliography, http://www.cs.uwyo.edu/~jhitchco/bib/dim.shtml (current April, 2011).

[15] J. M. Hitchcock. Resource Bounded Measure - Bibliography, http://www.cs.uwyo.edu/~jhitchco/bib/rbm.shtml (current April, 2011).

[16] J. M. Hitchcock. Fractal dimension and logarithmic loss unpredictability. *Theoretical Computer Science*, 304(1–3):431–441, 2003.

[17] John M. Hitchcock. Fractal dimension and logarithmic loss unpredictability, 2004.

[18] Chun-Yuan Hsiao, Chi-Jen Lu, and Leonid Reyzin. Conditional computational entropy, or toward separating pseudoentropy from compressibility. In *Advances in Cryptology - EUROCRYPT 2007, 26th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Barcelona, Spain, May 20-24, 2007, Proceedings*, pages 169–186, 2007.

[19] Russell Impagliazzo and Avi Wigderson. P=BPP unless E has sub-exponential circuits: Derandomizing the xor lemma (preliminary version). In *In Proceedings of the 29th STOC*, pages 220–229. ACM Press, 1996.

[20] Jesse Kamp and David Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. In *In Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 92–101, 2003.

[21] Chi-Jen Lu, Omer Reingold, Salil P. Vadhan, and Avi Wigderson. Extractors: optimal up to constant factors. In Lawrence L. Larmore and Michel X. Goemans, editors, *STOC*, pages 602–611. ACM, 2003.

[22] J. H. Lutz. Gales and the constructive dimension of individual sequences. In *Proceedings of the 27th International Colloquium on Automata, Languages, and Programming*, pages 902–913, 2000. Revised as [23].

[23] J. H. Lutz. The dimensions of individual strings and sequences. *Information and Computation*, 187:49–79, 2003. Preliminary version appeared as [22].

[24] Jack H. Lutz. Dimension in complexity classes. *SIAM J. Comput.*, 32(5):1236–1259, 2003.

[25] Jack H. Lutz. A divergence formula for randomness and dimension. *Theor. Comput. Sci.*, 412(1-2):166–177, 2011.

[26] Noam Nisan and Amnon Ta-Shma. Extracting randomness: A survey and new constructions. *J. Comput. Syst. Sci.*, 58(1):148–173, 1999.

[27] Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, 1994.

[28] Noam Nisan and David Zuckerman. More deterministic simulation in logspace. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing, May 16-18, 1993, San Diego, CA, USA*, pages 235–244, 1993.

[29] Noam Nisan and David Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52:43–52, 1996.

[30] Jaikumar Radhakrishnan and Amnon Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM Journal on Discrete Mathematics*, 13:2000, 2000.

[31] Omer Reingold, Luca Trevisan, and Salil Vadhan. Pseudorandom walks on regular digraphs and the rl vs. l problem. In *In Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC 06)*, pages 457–466, 2006.

[32] Ronen Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the EATCS*, 77:67–95, 2002.

[33] Luca Trevisan and Salil P. Vadhan. Extracting randomness from samplable distributions. In *FOCS*, pages 32–42. IEEE Computer Society, 2000.

[34] Salil P. Vadhan and Colin Jia Zheng. Characterizing pseudoentropy and simplifying pseudorandom generator constructions. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New*

York, NY, USA, May 19 - 22, 2012, pages 817–836, 2012.

[35] Heribert Vollmer. *Introduction to Circuit Complexity - A Uniform Approach*. Texts in Theoretical Computer Science. An EATCS Series. Springer, 1999.

[36] Andrew C. Yao. Theory and application of trapdoor functions. In *Proceedings of the 23rd Annual Symposium on Foundations of Computer Science*, SFCS '82, pages 80–91, Washington, DC, USA, 1982. IEEE Computer Society.