

## Lecture 12: Weak Sources of Randomness and Extractors

October 14, 2004

Scribe: Ehsan Amiri

## 1 Motivation

Probabilistic algorithms play a major role in computer science; for some problems, for example **IDENTITY TESTING**, there is no efficient deterministic algorithm, but instead there are quite efficient probabilistic ones. For some other problems like **PRIMALITY**, although we know the problem is in **P**, yet randomized algorithms are more efficient. Randomization is also a major ingredient in Cryptography, simply because in cryptography it is very important to behave in an "unpredictable" way, which means that some kind of randomization should be involved in the process (i.e. algorithms, protocols).

How we can get access to a source of randomness? How algorithms can toss a coin? In computer science, we usually need "truly random bits" which means a sequence of bits that each of them is 1 with probability  $\frac{1}{2}$  independent of other bits. Some physical processes, like radioactive decay, provide some kind of randomness but these sources of randomness do not satisfy our uniformity and independence conditions. So, it is desirable to be able to "extract" a sequence of "truly random bits" from these natural sources of "weak randomness". In this lecture and few next ones, we are going to formally define the above quoted words and show how we can do this task.

Once we can extract true randomness from weak sources of randomness, we will be able to build a "compiler" that helps us to simulate **BPP** algorithms in this way: Assume that  $A(., r)$  is an implementation of a **BPP** algorithm that uses a string  $r$  of truly random bits. Our compiler will generate a new program  $A'(., r')$  such that

- $|r'| \leq \text{poly}(|r|)$ .
- $r'$  is a sample taken from a weak random source.
- $A'$  is efficient if  $A$  is efficient.
- the error probability of  $A'$  is close to that of  $A$ .

## 2 Some Examples

**Example 1** *Assume that the source of randomness at our disposal is a biased coin. The result of tossing this coin would be 1 with probability  $\delta$  and 0 with probability  $1 - \delta$ . Toss this coin twice. If the result is 00 or 11 discard it, otherwise interpret 01 as 0 and 10 as 1. It's easy to see that this process simulates an unbiased coin and we need to toss the coin a constant number of times so it is also efficient. This kind of randomness sources are called Von Neumann's sources. Note that different samples of this source are independent. We will see that in general case we can not make such an assumption.*

**Example 2** Now we make a small change in the previous example. Assume that  $i$ -th flip has the probability  $\delta_i$  of being 1. Assuming that for some constant  $\delta > 0$  we have  $\delta \leq \delta_i \leq 1 - \delta$ , one can show that

$$\left| \Pr\left[\bigoplus_{i=1}^{\ell} X_i = 1\right] - \frac{1}{2} \right| = 2^{-\Omega(\delta\ell)}$$

**Exercises 3** Prove the above claim.

The sources of the following example are called Santha-Vazirani sources.

**Example 4** Now, we relax the independence assumption of the previous example. Assume that we can sample a sequence  $X_1, X_2, \dots, X_t$  of bits from a source of randomness such that for all  $1 \leq i \leq t$  and a constant  $\delta > 0$ , for any bit sequence  $b_1, \dots, b_{i-1}$ :

$$\delta \leq \Pr[X_i = 1 | X_1 = b_1, \dots, X_{i-1} = b_{i-1}] \leq 1 - \delta$$

One can ensure that the probability that the parity of the sequence  $X_1, X_2, \dots, X_t$  is 1 is at least  $\delta$  and at most  $1 - \delta$ . So the parity does not produce anything close to a fair coin flip. (More generally, one can prove that even one bit of randomness cannot be extracted from an SV-source.)

## 3 Measures of Randomness

### 3.1 Definitions

**Entropy** is one of the measures of randomness. There are different variations of entropy. Assume that  $X$  is a random variable over a fixed set  $U$ .

**Note:** all logarithms are base 2.

**Definition 5** Shannon's entropy of random variable  $X$  is denoted by  $\mathbf{H}_{\text{sh}}$  and defined as:

$$\mathbf{H}_{\text{sh}}(X) = \mathbf{E}_{x \in U} \left[ \log \frac{1}{\Pr[X = x]} \right]$$

**Definition 6** Renyi or  $\ell_2$  entropy is denoted by  $\mathbf{H}_2$  and defined as:

$$\mathbf{H}_2(X) = \log \frac{1}{\mathbf{E}_{x \in U} [\Pr[X = x]]} = \log \frac{1}{\mathbf{Col}(X)}$$

Where  $\mathbf{Col}(X)$  is the collision probability of  $X$  and defined in one of the previous lectures.

**Definition 7** Min-entropy or  $\ell_\infty$ -entropy is denoted by  $\mathbf{H}_\infty$  and defined as

$$\mathbf{H}_\infty(X) = \min_{x \in U} \left\{ \log \frac{1}{\Pr[X = x]} \right\} = \log \left\{ \frac{1}{\max_{x \in U} \{\Pr[X = x]\}} \right\}$$

It is easy to show that  $\mathbf{H}_\infty(X) \geq k$  implies that  $\forall x \in U : \Pr[X = x] \leq 2^{-k}$

### 3.2 Properties

In this subsection  $\mathbf{H}(\cdot)$  can be replaced by any of the above variations of entropy.

1. For the uniform distribution over  $\{0, 1\}^n$ , denoted by  $U_n$ ,  $\mathbf{H}(\mathbf{U}_n) = n$  which is the maximum possible value of all versions of entropy for any distribution over the set  $\{0, 1\}^n$ .
2.  $0 \leq \mathbf{H}(X) \leq \log |\mathbf{Supp}(X)|$ . The entropy is minimized if  $X$  is constant and maximized if distribution of  $X$  is uniform over the probability space.
3. For independent random variables  $X$  and  $Y$

$$\mathbf{H}(X, Y) = \mathbf{H}(X) + \mathbf{H}(Y)$$

where by  $(X, Y)$  we mean the joint distribution defined by  $X$  and  $Y$ .

4. For any function  $f$ ,  $\mathbf{H}(f(X)) \leq \mathbf{H}(X)$ . Intuitively the reason is clear, a function can not increase the randomness.
5. For all  $X$ 
  - $\mathbf{H}_\infty(X) \leq \mathbf{H}_2(X) \leq \mathbf{H}_{\text{sh}}(X)$
  - $\mathbf{H}_\infty(X) \leq \mathbf{H}_2(X) \leq 2\mathbf{H}_\infty(X)$

Here, we prove the second part of the last property. Suppose that  $P_{max} = \max_{x \in U} \mathbf{Pr}[X = x]$ .

$$P_{max}^2 \leq \sum_{x \in U} \mathbf{Pr}^2[X = x] \leq P_{max} \sum_{x \in U} \mathbf{Pr}[X = x] = P_{max}$$

But  $\sum_{x \in U} \mathbf{Pr}^2[X = x] = \mathbf{Col}(X)$ . So we get:

$$P_{max}^2 \leq \mathbf{Col}(X) \leq P_{max}$$

Using the second inequality of the above expression:

$$\mathbf{H}_2(X) = \log \frac{1}{\mathbf{Col}(X)} \geq \log \frac{1}{P_{max}} = \mathbf{H}_\infty(X)$$

and similarly the first inequality implies that  $\mathbf{H}_2(X) \leq 2\mathbf{H}_\infty(X)$ .

### 3.3 Shannon Entropy or Min Entropy?

**Remark 8** In this section we will use some simple facts about convex and concave function. A function  $f$  is convex in an interval  $[a, b]$  if for any two points  $c, d \in [a, b]$ :

$$f\left(\frac{c+d}{2}\right) \leq \frac{1}{2}(f(c) + f(d))$$

In the case that  $f$  has the second derivative in this interval the necessary and sufficient condition for convexity of  $f$  is that for any  $c \in [a, b]$ :  $f''(c) \geq 0$ . A function  $f$  is concave if and only if  $-f$  is convex.

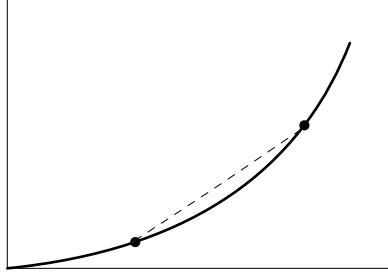


Figure 1: An example of a convex function

Here we make an example to show why Shannon entropy does not work for us. We will use **Jensen inequality** that says for any concave function and  $0 \leq \lambda \leq 1$ :

$$f(\lambda a + (1 - \lambda)b) \geq \lambda f(a) + (1 - \lambda)f(b)$$

Consider the following random variable  $X$

$$X = \begin{cases} 0^n & \text{with probability } 0.99 \\ U_n & \text{with probability } 0.01 \end{cases}$$

By the concavity of Shannon's entropy:

$$\begin{aligned} \mathbf{H}_{\text{sh}}(X) &= \mathbf{H}_{\text{sh}}(0.99C + 0.01U_n) & (1) \\ &\geq 0.99\mathbf{H}_{\text{sh}}(C) + 0.01\mathbf{H}_{\text{sh}}(U_n) = 0.01n & (2) \end{aligned}$$

Where, by  $C$  we mean the constant distribution that returns  $0^n$  with probability 1.

What does it mean? From the viewpoint of Shannon entropy the above random variable is a very good source of randomness because it provides  $\Omega(n)$  bits of randomness. In some sense Shannon is right: if one takes 100 samples of the above distribution, one has a good chance of taking at least one sample according to the uniform distribution and that very sample has lots of random bits. Recall that we want to take exactly one sample of the source and this sample should provide us with enough random bits while for the above source each single sample has a high chance of being all zero string, hence it does not provide any randomness.

Because of this difference between Shannon's entropy and our expectations we cannot use Shannon's entropy but it seems that the other two variations work better:

$$\begin{aligned} \mathbf{H}_2(X) &= \log \frac{1}{\mathbf{Col}(X)} \leq \log \frac{1}{0.99^2} < 1 \\ \mathbf{H}_\infty(X) &= \log \frac{1}{\max_{x \in \{0,1\}^n} \{\mathbf{Pr}(X = x)\}} \leq \frac{1}{0.99} < 1 \end{aligned}$$

As this example suggests Renyi and min entropy seems to fit our purpose. Henceforth we will mostly work with min-entropy.

## 4 $k$ -Sources

**Definition 9**  $X$  is a  $k$ -source if  $\mathbf{H}_\infty(X) \geq k$ ; i.e.  $\forall x, \Pr[X = x] \leq 2^{-k}$ .

**Example 10** A distribution which is uniform over its support is a  $k$ -source if the size of the support is at least  $2^k$ . These kind of distributions are called flat sources.

**Example 11** Consider a distribution over  $\{0, 1\}^n$  which has  $k$  random independent bits and  $n - k$  fixed bits. These sources are called oblivious bit fixing sources.

**Example 12** Adaptive bit fixing sources are distributions over  $\{0, 1\}^n$  that has  $k$  random fixed bits. The other  $n - k$  bits may arbitrarily depend on the random bits. These sources are also kind of  $k$ -sources.

**Example 13** Consider Santha-Vazirani sources with parameter  $\delta \leq \frac{1}{2}$ . For these sources we have  $\max_x \{\Pr[X = x]\} \leq (1 - \delta)^n \leq e^{-\delta n}$ . Consequently min-entropy of these sources is  $\theta(\delta n)$

The following interesting proposition makes it clear why  $k$ -sources are interesting for us.

**Proposition 14** Any  $k$ -source is a convex combination of flat sources.

Thanks to this proposition, to prove any claim about  $k$ -sources we only need to work on flat sources that have a simple structure. In the proof of the above proposition we use the following claim which is easy to prove:

**Claim 15** Any convex combination of two  $k$ -sources is a  $k$ -source.

**Exercises 16** Prove the claim.

**Proof Sketch:** One can consider a  $k$ -source over  $\{0, 1\}^n$  as a point in  $\mathfrak{R}^n$ . The above claim shows that the collection of all such sources is a convex polytope in  $\mathfrak{R}^n$ . Now using the fact that any point within or on a polytope is a convex combination of the vertices of the polytope we only need to show that the vertices of our polytope are exactly the points corresponding to the flat distributions.

To see this, note that faces of our polytope are planes  $X_i = 0$  and  $X_i = 2^{-k}$ . □

### 4.1 Extracting Randomness from $k$ -sources

Well, it's impossible!

**Theorem 17** For each function  $\mathcal{E} : \{0, 1\}^n \rightarrow \{0, 1\}$  there exists a flat  $(n - 1)$ -source  $X$  such that  $\mathcal{E}(X)$  is constant.

**Proof:** Clearly there exists  $b \in \{0, 1\}$  such that  $|\mathcal{E}^{-1}(b)| \geq \frac{2^n}{2} = 2^{n-1}$ . Take  $X$  to be uniform distribution on  $\mathcal{E}^{-1}(b)$  ■

So, is it possible to construct an extractor? The problem with the above definition is that function  $\mathcal{E}$  is deterministic. The behaviour of a deterministic function is predictable so we can fool it by constructing a bad example of a weak source. So there is still some hope that we can find a probabilistic algorithm that has the desired characteristics.

Before stating next theorem we need to define another measure of closeness of distributions:

**Definition 18** *Suppose that  $X$  and  $Y$  are two distributions over the same set  $U$ . **Statistical distance** of  $X$  and  $Y$  is*

$$\Delta(X, Y) = \frac{1}{2} \|X - Y\|_1 = \max_{T \subseteq U} |\Pr[X \in T] - \Pr[Y \in T]|$$

**Exercises 19** *Prove the second equality in the above definition.*

#### 4.1.1 Properties of statistical distance

1.  $0 \leq \Delta(X, Y) \leq 1$ .  $\Delta(X, Y)$  is minimized when  $X$  and  $Y$  are the same distribution and maximized when they have disjoint supports.
2.  $\Delta(X, Y) = \Delta(Y, X)$
3.  $\Delta(X, Z) \leq \Delta(X, Y) + \Delta(Y, Z)$
4. For any function  $f : U \rightarrow U$ ,  $\Delta(f(X), f(Y)) \leq \Delta(X, Y)$
5. For independent  $X_1, X_2$  and independent  $Y_1, Y_2$ :

$$\Delta([X_1, X_2], [Y_1, Y_2]) \leq \Delta(X_1, Y_1) + \Delta(X_2, Y_2)$$