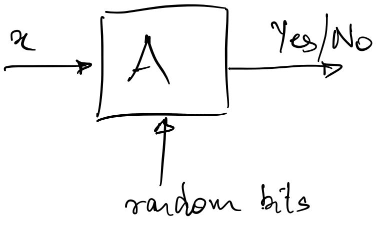
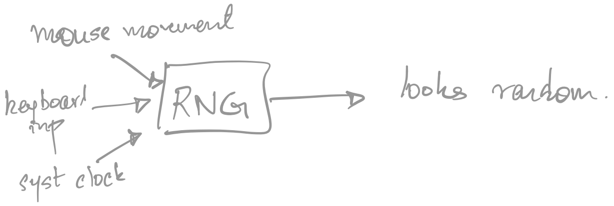
Pseudorandomness - Lecture 19.

Instructor: Ramprasad Saptharishi
Date: 2021-11-02
Lecture #: 19

Agenda: Introduction to extractors.



← uniformly random.



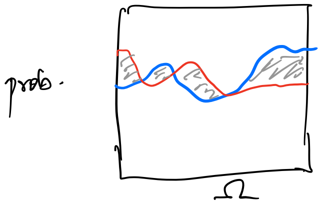
Extractors (informal): a way of "extracting" pure randomness from "impure" random sources.

$$\text{Ext}: X \longrightarrow \{0,1\}^m$$

↳ some "impure source".

Guarantee: - $\{\text{Ext}(X)\} \underset{TV}{\approx} \{\mathcal{U}_m\}$

Defn: $TV(X_1, X_2) = \frac{1}{2} \sum_{x \in \Omega} \left| \Pr[X_1 = x] - \Pr[X_2 = x] \right|$



Fact: If $TV(X,Y) \le \varepsilon$, then for any $A: \Omega \to \overset{\{0,1\}}{[-1,1]}$,

$$\left| \mathbb{E}[A(X)] - \mathbb{E}[A(Y)] \right| \le \varepsilon.$$

ie, X & Y are ε-close wrt any test A.

Fact: $TV(X,Y) = \inf_{\substack{\pi \sim \Omega^2 \\ \text{marginals} \\ X \,\&\, Y}} \Pr_{(x,y)\sim\pi}[x \neq y]$. $\pi = (\underset{X}{\downarrow}\,,\ \underset{Y}{\searrow})$

Bern(p) Bern(q)

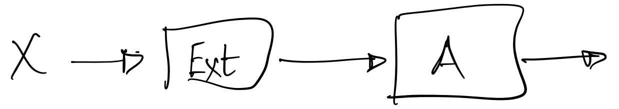
Some easy-to-verify properties:

- ▷ $1 \geq TV(X,Y) \geq 0$
- ▷ $TV(X,Y) + T(Y,Z) \geq TV(X,Z)$
- ▷ For any function f,
  $TV(f(X), f(Y)) \leq TV(X,Y)$.
- ▷ If X_1, X_2 are indep and so are Y_1, Y_2, then
  $TV((X_1,X_2),(Y_1,Y_2)) = TV(X_1,Y_1) + TV(X_2,Y_2)$

So that's our notion of "closeness".

Suppose we only have "impure" sources of randomness, can we still use them for randomised algos?

X - impure $Ext: X \to \{0,1\}^m$ A



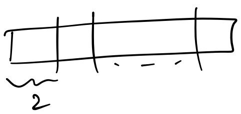
Differs by at most ε from the ideal setting.

What do "impure" sources mean?

Eg 1: IID-Bit$_\delta$ sources

$X_1, \dots, X_n \in \{0,1\}$. Each X_i i.i.d with $\Pr[X_i = 1] = \delta$.

Extractors? von Neuman:



01 → 0
10 → 1
00/11 → skip

→ output unbiased bits!

$\frac{1}{2\delta(1-\delta)} n \approx m$.

2. IndBits$_\delta$

$X_1, \dots, X_n \in \{0,1\}$, indep with $\delta \le \Pr[X_i = 1] \le 1-\delta$

Ext:

$\frac{1}{2} + (2\delta - 1)^k$. → close to uniform.

Prop: For any const. $\delta > 0$, and every $n, m \in \mathbb{N}$, there is a poly time fn. $\text{Ext} : \{0,1\}^n \to \{0,1\}^m$ that is an ε-extractor for IndBits$_\delta$ with $\varepsilon = m \cdot 2^{-\Omega(n/m)}$.

Important note: We only get a <u>single</u> sample of X and want to extract from it.

3. Unpredictable bit sources (Unpred.Bits$_\delta$)
(Santha-Vazirani sources).

$X_1, .., X_n \in \{0,1\}$. For every $i \in [n]$ and every $x_1, .., x_{i-1} \in \{0,1\}$,

$$\delta \leq \Pr[X_i = 1 \mid X_1 = x_1, .., X_{i-1} = x_{i-1}] \leq 1-\delta$$

(like next-bit-unpredictability we saw earlier)

Any candidates? Extract 1 close-to-random bit.

$$\mathrm{Ext}(X_1, .., X_n) = \bigoplus_{i=1}^{n} X_i$$

No! Set up s.t $X_n = \bigoplus_{i=1}^{n-1} X_i ..$ w.p $1-\delta$

$\mathrm{Maj}(X_1, . X_n)$

Prop (in PS4): For any $n \in \mathbb{N}$, $\delta > 0$ and <u>any</u> fn $\mathrm{Ext}: \{0,1\}^n \to \{0,1\}$, there is an UnpredBits$_\delta$ source X s.t

$$\Pr[\mathrm{Ext}(X) = 1] \leq \delta \text{ or } \Pr[\mathrm{Ext}(X) = 1] \geq 1-\delta.$$

"There are <u>no</u> <u>det</u>. extractors"

Quantitative measures for weak/impure sources.

Defn: (Shannon entropy).

$$H_{Sh}(X) = \sum_{x\in\Omega} p_x \log\left(\frac{1}{p_x}\right) = \mathop{\mathbb{E}}_{X}\left[\log\left(\frac{1}{p_x}\right)\right]$$

Essential in the field of communication theory & information theory, but meant to understand "behavior on average". (asymptotics)

Defn (Rényi entropy):

$$H_2(X) = \log\left(\frac{1}{\Pr_{x_1,x_2\sim X\times X}[x_1 = x_2]}\right)$$

$$= \log\left(\frac{1}{CP(X)}\right)$$

Defn: (Min-entropy)

$$H_\infty(X) = \min_{x : p_x \neq 0}\left\{\log \frac{1}{p_x}\right\}$$

$H_\infty(X) = k$

$\Rightarrow \Pr[X = x] \leq 2^{-k}$
$\forall x$.

Basic properties: ▷ $H(X) \geq 0$, eq only when X has singleton supp.

▷ If X = uniform on a subset of size 2^k, then
$H(X) = k$.

▷ X, Y indep $\Rightarrow H(X,Y) = H(X) + H(Y)$.

$H_\infty(X) \leq H_2(X) \leq H_{Sh}(X)$.

$$X = \begin{cases} 0^n & \text{w.p} \quad 0.99 \\ \text{uniform} & \text{w.p} \quad 0.01 \end{cases}$$

$H_{Sh}(X) \approx 0.01\, n$.

$\Big\downarrow$

How can I get even 1 bit of randomness from a single sample?!

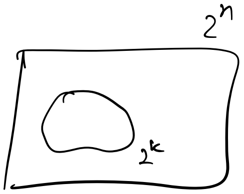
$H_\infty(X) \approx \log\left(\frac{1}{0.99}\right)$
≤ 2.

$H_2(X) \approx \log\left(\frac{1}{(0.99)^2}\right)$

We'll mostly work with min-entropy.

Defn: (Weak k-source) A RV X is a k-source if $H_\infty(X) \geq k$ (or $\Pr[X = x] \leq 2^{-k}$ for any x).

All examples earlier have $H_\infty(X) = \Omega_\delta(n)$



Other examples:

- Bit-fixing sources: $\boxed{010*011*01*00*10}$
  Some k bits are truly uniform, rest are always fixed.
- Adaptive bit fixing sources.
  k-bits uniform. Rest are some specific fn of these k.
- Flat k-sources: uniform dist on some set of size exactly 2^k

Fact: Every k-source is a conv. comb of flat k-sources.

Why are we doing all this when $\boxed{\text{det}}$ extractors don't exist?

Prop: For any fn: $\text{Ext}: \{0,1\}^n \to \{0,1\}$, there is an $(n-1)$ source on which $\text{Ext}(X)$ is constant.

Pf: $X = \text{Ext}^{-1}(0)$ or $\text{Ext}^{-1}(1)$ (whichever is larger) $\square$.

Det. extractors don't exist...

Defn: (Seeded extractors) A function $\text{Ext}: \{0,1\}^n \times \overset{\text{truly uniform.}}{\{0,1\}^s} \to \{0,1\}^m$ is said to be a (k,ε)-seeded extractor if for every k-source X, we have $\text{Ext}(X, \mathcal{U}_s) \overset{TV}{\approx}_{\varepsilon} \mathcal{U}_m$.

Ideally want $m \approx k+s.$, efficient. and for k as small as possible.

Do such extractors exist?

Lemma: For all n, k and $\varepsilon > 0$, if $s = \log(n-k) + 2\log\frac{1}{\varepsilon} + O(1)$, and $m = k + s - 2\log\frac{1}{\varepsilon} - O(1)$, a random function $\text{Ext}: \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$ is a (k,ε)-extractor w.h.p.

Pf: Suffices to show it works for flat k-sources.

Fix a source X.

$\text{Ext. fails} \Rightarrow \exists\, T \subset \{0,1\}^m$ s.t

$2^k\, 2^s$

$$\left| \Pr\left[\text{Ext}(\underbrace{X, \mathcal{U}_s}) \in T \right] - \frac{|T|}{2^m} \right| > \varepsilon$$

$\text{Ext}: \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$

$$\Rightarrow \Pr_{\text{ext}}\left[\text{Ext fails on } X \text{ wrt } T \right] \leq 2^{-\Omega(2^{k+s} \cdot \varepsilon^2)}$$

$$\Rightarrow \Pr_{\text{Ext}}\left[\text{Ext fails for some } X, T\right] \leq \binom{2^n}{2^k} \cdot 2^{2^m} \cdot 2^{-\Omega(2^{k+s}\varepsilon^2)}$$

" Wave hands "
for the choice of parameters in the statement, RHS << 1.

$\Rightarrow$ Ext is indeed an extractor w.h.p. □.

∴ Seeded extractors, with $O(\log n)$ seed-length <u>do</u> exist.

Prop: If such explicit efficient seeded extractors exist, then any randomised algo can be simulated on weak sources

Pf: X - weak k-source

A - rand. algo that needed m pure random bits.

diff is at most ε. ← efficient simulation.

$\text{Ext}: \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$
(k, ε)- extractor, efficient

▷ Pick $x \sim X$

▷ $\underset{y \sim \{0,1\}^s}{\text{maj}} \{A(\text{Ext}(x,y))\}$

Focus of next few lectures: constructing such explicit, efficient extractors.