RANDOMNESS EXTRACTION

Recall: No $Ext : \{0,1\}^n \to \{0,1\}^\ell$ can output s.t. UNIFORM even if input X is s.t. $H_\infty(X) = n-1$

Way out: Let $Ext : \{0,1\}^d \times \{0,1\}^n \to \{0,1\}^\ell$ where the first input is UNIFORM SEED $S \equiv V_d$

Why non trivial:

- $\ell >> d$
- Seed S public (aka "REUSABLE").

To make it easier: Allow Ext some error ε, meaning output of Ext only "close" to UNIFORM.

DEF. (SD) The SD between X, X' over $\mathcal{X}$ is

$$SD(X; X') = \frac{1}{2} \sum_{x \in \mathcal{X}} \left| \Pr[X = x] - \Pr[X' = x] \right|$$

FACT It's a metric.

Equivalent to say: No UNBOUNDED A can distinguish $x \leftarrow\!\$\, X$ from $x \leftarrow\!\$\, X'$ (w.p. $\geq \varepsilon$)

DEF (SEEDED EXT) $Ext : \{0,1\}^d \times \{0,1\}^n \to \{0,1\}^\ell$

is a (k, ε)-extractor if: $\forall X \in \{0,1\}^n$ s.t. $H_\infty(X) \geq k$

$$SD\left(\left(S, \underbrace{Ext(S,X)}_{Y}\right), \left(S, U_\ell\right)\right) \leq \varepsilon$$

where $S \equiv U_d$.

THM (LEFTOVER HASH LEMMA, HILL '89) Let $\mathcal{H} = \{h_s : \{0,1\}^n \to \{0,1\}^\ell\}_{s \in \{0,1\}^d}$ be PI. Then $Ext(s,x) = h_s(x)$ is a (k, ε)-ext. for $k \geq \ell + 2\log(1/\varepsilon) - 2$.

LEMMA. Let Y be RV over $\mathcal{Y}$ s.t.

$$\mathrm{Col}(Y) = \sum_{y \in \mathcal{Y}} \Pr[Y = y]^2$$

$$\leq \frac{1}{|\mathcal{Y}|} \cdot (1 + 4\varepsilon^2).$$

$$\Rightarrow \mathrm{SD}(Y, U) \leq \varepsilon$$

↳ over $\mathcal{Y}$.

Proof (of THM). Take $Y = (S, h(S,X))$

$$Y' = (S', h(S',X'))$$

Now show $Col(Y) \leq \frac{1}{|y|} \cdot (1 + 4\varepsilon^2)$.

$$Col(Y) = Pr[Y = Y']$$

$$= Pr[S = S' \wedge h(S,X) = h(S',X')]$$

$$= Pr[S = S' \wedge h(S,X) = h(S,X')] \quad (\text{as } S = S')$$

$$= \underbrace{Pr[S = S']}_{Col(U_d)} \cdot Pr[h(S,X) = h(S,X')]$$

$$= 2^{-d} \cdot \left(\Pr[X = X'] + \Pr[h(S,X) = h(S,X') \wedge X \neq X'] \right)$$

$$\leq 2^{-d} \cdot \left(2^{-k} + \Pr[h(S,R) = h(S,X') \wedge X \neq X'] \right)$$

$$= 2^{-d} \cdot \left(2^{-k} + 2^{-\ell} \right) \qquad \begin{array}{l} (\text{because } H_\infty(X) \geq k) \\ (\text{by } P1) \end{array}$$

$$= \frac{1}{2^{d+\ell}} \left(2^{\ell-k} + 1 \right)$$

$$\leq \frac{1}{|\mathcal{Y}|} \left(2^{2-2\log(1/\varepsilon)} + 1 \right) = \frac{1}{|\mathcal{Y}|} \cdot (4\varepsilon^2 + 1) \quad \square$$

Proof (of LEMMA) By def:

$$SD(Y;U) = \frac{1}{2}\sum_{y\in\mathcal{Y}} \left| \underbrace{\Pr[Y=y] - \frac{1}{|\mathcal{Y}|}}_{q_y} \right|$$

$$q_y = \Pr[Y=y] - 1/|\mathcal{Y}|$$

$$s_y = \begin{cases} 1 & \text{if } q_y \geq 0 \\ -1 & \text{if } q_y < 0 \end{cases}$$

$$SD(Y;U) = \frac{1}{2}\sum_{y} q_y s_y = \frac{1}{2} \langle \vec{q}, \vec{s} \rangle$$

$$\leq \frac{1}{2} \sqrt{\langle \vec{q}, \vec{q} \rangle \cdot \langle \vec{s}, \vec{s} \rangle} \quad (CS)$$

$$= \frac{1}{2} \sqrt{\sum_y q_y^2 \cdot |y|}$$

$$\left(\langle \vec{q}, \vec{q} \rangle = \sum_y q_y^2 \; ; \langle \vec{s}, \vec{s} \rangle = |y| \right)$$

$$\Rightarrow \boxed{SD(Y;U) \leq \frac{1}{2} \sqrt{\sum_y q_y^2 \cdot |y|}}$$

$$\sum_y q_y^2 = \sum_{y=y} \left(P_r[Y=y]^2 - \frac{2\, P_r[Y=y]}{|y|} + \frac{1}{|y|^2} \right)$$

$$= Col(Y) - \frac{2}{|y|} + \frac{1}{|y|}$$

$$\leq \frac{1}{|y|}\left(4\varepsilon^2+1\right)\frac{1}{4} - \frac{1}{|y|}$$

$$= 4\varepsilon^2/|y|$$

$$\Rightarrow SD(Y_1 \cdot U) \leq \frac{1}{2}\sqrt{4\varepsilon^2} = \varepsilon \quad \square$$