3 Randomness extraction

3.1 Problem setup

$P_X(0) = P_X(1) = \frac{1}{2}$

${}^a X = 101011001$

${}^b X = 010000000$

$\Pr[X^9 = {}^a X] = 2^{-9}$

$\Pr[X^9 = {}^b X] = 2^{-9}$

$$\delta_{tvd}(P_Z, U_Z) := \frac{1}{2}\|P_Z - U_Z\|_1 = \frac{1}{2}\sum_{z \in \mathcal{Z}} |P_Z(z) - U_Z(z)|$$

measures the distance to uniform

$U_Z(z) = \frac{1}{|\mathcal{Z}|}$ for all $z \in \mathcal{Z}$

uniform distribution

Data-processing inequality:

$$\delta_{tvd}(P_X, Q_X) \geq \delta_{tvd}(P_Y, Q_Y)$$

where $P(y) = \sum_x W(y|x)P(x)$, $Q(y) = \sum_x W(y|x)Q(x)$

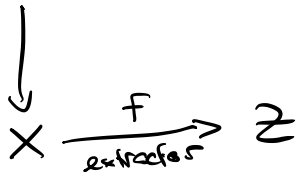
When is P_{XV} secret (that is, X is secret from V)?

Def: Let P_{ZV} be a joint distribution. For any $\varepsilon \in [0,1]$ we say that Z is ε-random and independent of V (or ε-secret) if

$$\delta_{tvd}(P_{ZV}, U_Z \times P_V) \leq \varepsilon.$$

side information

Source ——> V



3.2 Guessing probability

P_{XY} is known. A pair (x,y) is drawn. If you know y, what is your best guess of x?

$$\hat{x} = \arg\max_x P(x|y)$$

$$P_{guess}(X|Y) = \sum_y P(y) \max_x P(x|y)$$

$$H_{min}(X|Y) = -\log P_{guess}(X|Y)$$

Example: $P_X(1) = \frac{1}{2}$

$P_X(2) = P_X(3) \dots P(10) = \frac{1}{18}$

$\Rightarrow P_{guess}(X) = \frac{1}{2}$

$H_{min}(X) = 1$

$Z = f(X)$

$$f(x) = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{else} \end{cases}$$

gives perfect uniformly random bit

$$-\log \max_x P(x) = \min_x \log \frac{1}{P(x)}$$

minimum surprisal

Lemma: $H_{min}(X|Y) \leq H(X|Y)$

Proof: $H_{min}(X|Y) = -\log \left(\sum_y P(y) \max_x P(x|y) \right)$

$$\leq \sum_y P(y) \left(-\log \max_x P(x|y) \right)$$

$$\min_x -\log P(x|y)$$

$$\leq \sum_y P(y) \sum_x P(x|y) \left(-\log P(x|y) \right)$$

$$= H(X|Y)$$

Def. ε-smooth min-entropy of P_{XY} is defined as

$$H_{min}^{\varepsilon}(X|Y)_P := \max \{ H_{min}(X|Y)_{\tilde{P}} : \delta_{tvd}(P_{XY}, \tilde{P}_{XY}) \leq \varepsilon \}$$

Def. An (ε, ℓ)-extractor for P_{XY} is a function

$F: X \to \{0,1\}^{\ell}$ such that

$\delta_{tvd}(P_{ZY}, U_Z \times P_Y)$ where $Z = f(X)$

and $P_{ZY}(z,y) = \sum_{x: f(x)=z} P_{XY}(x,y)$.

$\ell^*_\varepsilon(X|Y) = \max \{ \ell : \exists\, (\varepsilon,\ell) \text{ extractor} \}$

we want to find upper and lower bounds on $\ell^*_\varepsilon(X|Y)$.

3.3 Achievability via two-universal hash functions

Family of functions $f_s : \mathcal{X} \to \mathcal{Z}$ ($= \{0,1\}^\ell$) where s is the seed. A two-universal family has the property that

$$\sum_{s \in S} \frac{1}{|S|} \mathbb{1}\{f_s(x) = f_s(x')\}$$

$$= \Pr[f_S(x) = f_S(x')] \leq \frac{1}{2^\ell} \quad \forall x \neq x'$$

where S is drawn uniformly at random

Theorem: Let $\{f_s\}$ be a two-universal family.

Then $\leq \frac{1}{...} S_{...}(P^S_{ZY}, (U_Z \times P_Y))$

then $\sum_{s \in S} |S|$ dtvd$(\cdot, \cdot)$ $\leq \frac{1}{2}\sqrt{2^{\ell - H_{\min}(X|Y)}}$

where $P^s_{ZY}(z,y) = \sum_{x : f_s(x) = z} P_{XY}(x,y)$.

Proof: $2\,\delta_{tvd}(P^s_{ZY}, U_Z \times P_Y)$

$= \| P^s_{ZY} - U_Z \times P_Y \|_1$

$= \| (1_Z \times P_Y^{1/2}) \cdot (1_Z \times P_Y^{-1/2}) \cdot (P^s_{ZY} - U_Z \times P_Y) \|_1$ (all 1 vector; element-wise multiplication)

$\leq \| 1_Z \times P_Y^{1/2} \|_2 \, \| (1_Z \times P_Y^{-1/2})(P^s_{ZY} - U_Z \times P_Y) \|_2$

$= \sqrt{\sum_{z,y} P_Y(y)} \cdot \sqrt{\sum_{z,y} P_Y(y)^{-1} \left(P^s_{ZY}(z,y) - \frac{1}{2^\ell} P_Y(y) \right)^2}$

$\sqrt{\sum_{z,y} P_Y(y)} = \sqrt{2^\ell}$

$= \sqrt{2^\ell \sum_{z,y} \left(P_Y(y)^{-1} P^s_{ZY}(z,y)^2 - \frac{2}{2^\ell} P^s_{ZY}(z,y) + \frac{1}{2^{2\ell}} P_Y(y) \right)}$

$= \sqrt{2^\ell \sum_{z,y} P_Y(y)^{-1} P^s_{ZY}(z,y)^2 - 1}$

$\Rightarrow$

$\sum \frac{1}{|S|} \delta_{tvd}(P^s_{ZY}, U_Z \times P_Y)$

$s \in S$ ...

$$\leq \frac{1}{2}\sqrt{2^\ell \sum_s \frac{1}{|S|} \sum_{z,y} P_Y(y)^{-1} P_{Z|Y}^s(z,y)^2 - 1}$$

$$\sum_s \frac{2^\ell}{|S|} \sum_{z,y} \sum_{x,x'} P_Y(y)^{-1} P_{X|Y}(x,y) P_{X|Y}(x',y) \mathbb{1}\{f_s(x) = z\} \mathbb{1}\{f_s(x') = z\}$$

$$= \sum_s \frac{2^\ell}{|S|} \sum_{y,x,x'} P_Y(y) \, P_{X|Y}(x|y) P_{X|Y}(x'|y) \, \mathbb{1}\{f_s(x) = f_s(x')\}$$

split into $x \neq x'$ and $x = x'$

$$= 2^\ell \sum_{\substack{x,x',y \\ x \neq x'}} P_Y(y) P_{X|Y}(x|y) P_{X|Y}(x'|y) \cdot \underbrace{\sum_s \frac{1}{|S|} \mathbb{1}\{f_s(x) = f_s(x')\}}_{\leq 2^{-\ell}}$$

$$+ 2^\ell \sum_{x,y} P_Y(y) P_{X|Y}(x|y)^2$$

$$\leq 1 + 2^\ell \sum_y P_Y(y) \underbrace{\sum_x P_{X|Y}(x|y) \cdot P_{X|Y}(x|y)}_{\leq \max_x P_{X|Y}(x|y)}$$

$$= 1 + 2^\ell P_{\text{guess}}(X|Y)$$

$= 1 + 2^{\ell - H_{min}(X|Y)}$ □

Theorem: Consider $\epsilon \in (0,1)$ and a source P_{XY}.
There exists an (ϵ, ℓ) extractor if

$$\ell \leq H_{min}^{\epsilon/4}(X|Y)_P - 2\log\frac{1}{\epsilon}.$$

Proof: Skipped

3.4 Converse via an entropy inequality

Lemma: $H(X) \geq H(f(X))$ (1), $H_{min}(X) \geq H_{min}(f(X))$ (2)
For any function F.

Proof: $Z = f(X)$, use joint distribution

$$P_{XZ}(x,z) = P_X(x)\,\mathbb{1}\{f(x) = z\}$$

1) $$H(XZ) = \sum_{x,z} P(x,z)\log\frac{1}{P(x,z)}$$

$$= \sum_x P(x)\log\frac{1}{P(x)} = H(X)$$

$$H(X) = H(XZ) = H(Z) + H(X|Z) \geq H(Z)$$

2) to show: $p_{guess}(X) \leq p_{guess}(Z)$

$$\max_x P(x) \leq \max_z \sum_{x:\, f(x)=z} P(x)$$

Lemma: Let $\varepsilon \in [0,1)$ and $f: \mathcal{X} \to \mathcal{Z}$ a function.

Then $H_{\min}^{\varepsilon}(X|Y) \geq H_{\min}^{\varepsilon}(f(X)|Y)$.

Proof: $\omega_{Z|X}(z|x) = \delta_{z,f(x)} = 1\{z = f(x)\}$

$$\tilde{\omega}_{X|ZY} = \frac{P_{XZ|Y}(x,z|y)}{P_{Z|Y}(z,y)} = \frac{\delta_{z,f(x)} P_{X|Y}(x|y)}{\sum_{\substack{x': \\ f(x')=z}} P_{X|Y}(x|y)}$$

This is a right-inverse of $\omega_{Z|X}$:

For any Q_{ZY}:

$$\tilde{Q}_{ZY}(z,y) = \sum_{x'} \omega_{Z|X}(z|x) \sum_{z'} \tilde{\omega}_{X|ZY}(x'|z',y) Q_{ZY}(z',y)$$

$$= Q_{ZY}(z,y)$$

Let us now take Q_{ZY} such that

$$H_{\min}^{\varepsilon}(Z|Y)_P = H_{\min}(Z|Y)_Q.$$

Construct $Q_{XY}(x,y) = \sum_{z'} \tilde{\omega}_{X|YZ}(x|z',y) Q_{ZY}(z',y)$

By data-processing, we have $\delta_{\text{tvd}}(Q_{XY}, P_{XY}) \leq \delta_{\text{tvd}}(Q_{ZY}, P_{ZY}) \leq \varepsilon$

$$H_{\min}^{\varepsilon}(X|Y)_P \geq H_{\min}(X|Y)_Q = -\log P_{\text{guess}}(X|Y)_Q$$
$$\geq H_{\min}(Z|Y)_Q = H_{\min}^{\varepsilon}(Z|Y)_P$$

$$p_{guess}(X|T)_Q = \sum_t Q_T(t)\, p_{guess}(X)_{Q_{X|T=t}}$$

$$\leq \sum_t Q_T(t)\, p_{guess}(Z)_{Q_{Z|T=t}}$$

$$= p_{guess}(Z|T)_Q$$

□

Theorem: Consider $\varepsilon \in (0,1)$. Then any (ε, ℓ)-extractor must satisfy

$$\ell \leq H_{min}^{\varepsilon}(X|T)_P$$

or, $\ell_\varepsilon^*(X|T) \leq H_{min}^{\varepsilon}(X|T)_P$

Proof: If f is (ε, ℓ)-extractor, then

$$\delta_{tvd}(P_{ZT}, U_Z \times P_T) \leq \varepsilon \text{ by definition}$$

$$\Rightarrow H_{min}^{\varepsilon}(Z|T)_P \geq H_{min}(Z|T)_{U \times P}$$

$$= \ell$$

Because of the previous lemma:

$$H_{min}^{\varepsilon}(X|T) \geq H_{min}^{\varepsilon}(Z|T) \geq \ell$$

□

Example

P_{XY}

	0	1	2
0	0	$\frac{1}{6}$	$\frac{1}{6}$
1	$\frac{1}{6}$	0	$\frac{1}{6}$
2	$\frac{1}{6}$	$\frac{1}{6}$	0

$P_X(0) = P_X(1) = P_X(2) = \frac{1}{3}$

$P_{guess}(X|Y) = \frac{1}{2}$