3
Application to cryptography: Randomness extraction

Intended learning outcomes:

- You can compute and use guessing probability and min-entropy.
- You can construct a randomness extractor using a family of hash functions.
- You understand that deterministic functions cannot increase entropy.

3.1 Randomness and problem setup

One of the most prominent concepts in cryptography is randomness, and it lies at the core of information-theoretic security. To understand, for example, whether a given bit string is *random*, we do not want to look at a particular instance of the string (although that is interesting as well and leads ultimately to the notion of algorithmic randomness) but instead want to check that the process that created the bit string selected it at random. Similarly and maybe even more evidently, the concept of a *secret* bit string cannot be defined unless we look at the process by which the bit string is produced. If the random process is such that the bit string is independent of any side information held by an eavesdropper, then secrecy (relative to that eavesdropper) can be claimed.

3.1.1 Total variation distance

Before we can discuss randomness and privacy we need to introduce the total variation distance.

The *total variation distance* (TVD) between two pmfs P_0 and P_1 is

$$\delta_{\mathrm{tvd}}(P_0, P_1) := \frac{1}{2} \sum_{x \in \mathcal{X}} \left| P_0(x) - P_1(x) \right| . \tag{3.1}$$

The total variation distance vanishes if and only if $P_0 = P_1$ and it reaches its maximum 1 when P_0 and P_1 are orthogonal, that is, when

for every $x \in \mathcal{X}$ either $P_0(x) = 0$ or $P_1(x) = 0$. We can alternatively express the TVD using the following variational formulae, which motivate its name.

Lemma 3.1. *For any two pmfs P_0 and P_1, the following relations hold:*

$$\delta_{\text{tvd}}(P_0, P_1) = \max_{\mathcal{A} \subset \mathcal{X}} \left(\sum_{x \in \mathcal{A}} P_0(x) - P_1(x) \right) \tag{3.2}$$

$$= \max_{\mathcal{A} \subset \mathcal{X}} \left(\sum_{x \in \mathcal{A}} P_1(x) - P_0(x) \right). \tag{3.3}$$

Proof. To see this equivalence, first note that

$$\sum_{x \in \mathcal{X}} P_0(x) - P_1(x) = 0 \tag{3.4}$$

by normalisation, and thus, for any set $\mathcal{A} \subseteq \mathcal{X}$, we have

$$\sum_{x \in \mathcal{A}} P_0(x) - P_1(x) = \sum_{x \in \mathcal{A}^c} P_1(x) - P_0(x). \tag{3.5}$$

Specifically, for the set $\mathcal{A} = \{x \in \mathcal{X} : P_0(x) > P_1(x)\}$ that is optimal for the maximisation in Eq. (3.2), we find

$$\sum_{x \in \mathcal{A}} \left| P_0(x) - P_1(x) \right| = \sum_{x \in \mathcal{A}^c} \left| P_1(x) - P_0(x) \right| \tag{3.6}$$

and thus

$$\max_{\mathcal{A} \subset \mathcal{X}} \left(\sum_{x \in \mathcal{A}} P_0(x) - P_1(x) \right) = \sum_{x \in \mathcal{A}} \left| P_0(x) - P_1(x) \right| \tag{3.7}$$

$$= \frac{1}{2} \sum_{x \in \mathcal{X}} \left| P_0(x) - P_1(x) \right|. \tag{3.8}$$

The second equality can be verified similarly. □

The total variational distance is closely related to the 1-norm distance, which is defined for any vectors v_0 and v_1 that do not necessary need to be normalised. Recall its definition from Section 0.4:

$$\|v_0 - v_1\|_1 = \sum_{x \in \mathcal{X}} \left| v_0(x) - v_1(x) \right|. \tag{3.9}$$

Hence, in particular, $\delta_{\text{tvd}}(P_0, P_1) = \frac{1}{2}\|P_0 - P_1\|_1$.

The total variation distance satisfies the DPI.

Proposition 3.2 (DPI for TVD)**.** *For any channel $W_{Y|X}$ and any two pmfs P_X and Q_X, we have*

$$\delta_{\text{tvd}}(P_X, Q_X) \geq \delta_{\text{tvd}}(P_Y, Q_Y), \tag{3.10}$$

Verify that the total variation distance is a metric: it is symmetric, it is positive and zero only if the two distributions are equal, and it satisfies the triangle inequality.

where the output distribution are given as in Prop. 1.9.

This can be understood as saying that after we apply a channel $W_{Y|X}$, that is, introduce some noise, the output distributions are generally closer than the input distributions. So in a sense the two distributions have become more difficult to distinguish after applying the channel. We will verify this property in the homework.

In the next chapter we learn more about the total variation distance and its use in statistics.

3.1.2 *Randomness*

In the following we say that a random variable Z on an alphabet $\mathcal{Z}$ is close to uniformly random if it pmf is close to a uniform pmf in TVD, i.e. if

$$\delta_{\text{tvd}}(P_Z, U_Z) = \frac{1}{2} \sum_{z \in \mathcal{Z}} |P_Z(z) - U_Z(z)| \tag{3.11}$$

is small, where U_Z denotes the uniform distribution on $\mathcal{Z}$.

We can now extend the definition of uniformity to the case where some side information Y on Z is available, and we want to make sure that the randomness is in fact not only uniform but also independent of the side information. This leads us to the following more general definition.

Let P_{ZY} be a joint pmf for two random variables Z on $\mathcal{Z}$ and Y on $\mathcal{Y}$. For any $\epsilon \in (0,1)$, we say that Z is *ϵ-uniformly random and independent of Y* if

$$\delta_{\text{tvd}}(P_{ZY}, U_Z \times P_Y) \leq \epsilon\,. \tag{3.12}$$

Here, it is worth noting that the TVD can be simplified to

$$\delta_{\text{tvd}}(P_{ZY}, U_Z \times P_Y) = \frac{1}{2} \sum_y P_Y(y) \sum_z |P(z|y) - U(z)|\,, \tag{3.13}$$

and thus what we really require is that $P(z|y)$ is close to uniform in expectation over y.

3.1.3 *Randomness extractors*

We will now consider the task of *randomness extraction*, namely the task of creating approximately uniform and independent random variables from a random source X that is generally neither uniform nor independent of Y. In cryptography the i.i.d. assumption (as appears for example in memoryless sources) is often not very natural since we often cannot guarantee that a random source is exactly

memoryless. We want to ensure that our extraction scheme works even if we do not make any assumptions on the structure of the source. See Figure 3.1 for a schematic.

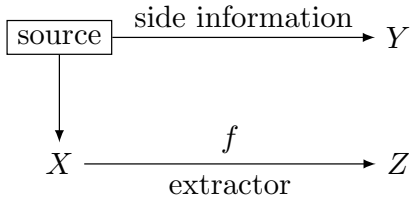

Figure 3.1: The setup of randomness extraction. A source produces random variables X and Y, where the latter variable Y is considered as side information on X. An extractor f is used to create a new random variable Z that is (close to) uniform and independent of Y. An important special case occurs when Y is trivial and we do not have any side information.

This is generally difficult: one thing we can immediately notice is that if one output of the source is very likely, for example if it appears exactly with probability 0.5, then we can produce exactly one bit of perfect randomness from this source (the new uniform random variable would be the indicator function for this event, which takes the value 0 and 1 with probability 0.5 each.), and this is in fact the best we can hope for. The maximal probability over any output of the source thus appears prominently in the analysis of randomness extraction, even in the approximate case, and we will introduce it formally in the next section in terms of guessing probability and min-entropy.

Let us now formally define a randomness extractor for a fixed source, which takes X and produces a bit string Z that is uniformly random and independent of Y.

An $(\epsilon, 2^L)$-extractor for a source X with side information Y governed by a pmf P_{XY} is a function $f : \mathcal{X} \to \{0,1\}^L$ such that

$$\delta_{\text{tvd}}(P_{ZY}, U_Z \times P_Y) \leq \epsilon \qquad \text{where} \qquad Z = f(X) \tag{3.14}$$

and thus P_{ZY} is the distribution induced by f, i.e.

$$P_{ZY}(z,y) = \sum_{x:f(x)=z} P_{XY}(x,y)\,. \tag{3.15}$$

We may now ask for the maximum length L of such an approximately uniform and independent string of bits. For this purpose we define

$$L^*_\epsilon(X|Y)_P := \max\left\{L \in \mathbb{N} : \exists \text{ an } (\epsilon, 2^L)\text{-extractor for } P_{XY}\right\}. \tag{3.16}$$

We will now find bounds on this quantity from above and below that hold for arbitrary distributions P_{XY}. These bounds will be in terms of the smooth min-entropy of the source, which we will introduce in the next section. In the homework we will also consider the special case where these sources are memoryless.

Example. *Consider a joint distribution P_{XY} given by the following table:*

P_{XY}	$x=1$	$x=2$	$x=3$
$y=1$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{6}$
$y=2$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{12}$

Clearly Y contains information about X; we can in fact compute $H(X) = \frac{3}{2}$ and

$$I(X:Y) = \frac{3}{2} - \frac{2}{3} - \frac{1}{2}\log 3 > 0\,.$$

Nonetheless, there is a strategy f to extract one perfect secret bit from X by mapping $\{1,3\} \mapsto 0$ and $2 \mapsto 1$. Then, for $Z = f(X)$ we find

$$P_{Z|Y}(\cdot|1) = P_{Z|Y}(\cdot|2) = \left(\frac{1}{2}, \frac{1}{2}\right).$$

Hence, Z is not correlated to Y.

Verify that the above extractor works by checking the TVD condition.

Why should we not allow random functions/channels as extractors here?

3.1.4 *Guessing probability and min-entropy*

We again consider a joint pmf P_{XY} on two random variables X on $\mathcal{X}$ and Y on $\mathcal{Y}$. We characterise our source using the concepts of guessing probability and min-entropy. The guessing probability of X given Y is the probability that an observer with access to Y can correctly guess the value of X. It is not difficult to find the optimal strategy for this task: given a sample $y \in \mathcal{Y}$, the observer will simply choose its guess as

$$\hat{x} = \operatorname*{argmax}_{x \in \mathcal{X}} P_{X|Y}(x|y)\,. \tag{3.17}$$

The average probability of guessing the correct value of X is thus given by the guessing probability as defined in the following.

Let P_{XY} be a joint pmf as above. The *guessing probability* of X conditioned on Y is defined as

$$p_{\text{guess}}(X|Y)_P := \sum_{y \in \mathcal{Y}} P_Y(y) \max_{x \in \mathcal{X}} P_{X|Y}(x|y)\,. \tag{3.18}$$

Moreover, the *conditional min-entropy* of X conditioned on Y is

$$H_{\min}(X|Y)_P := -\log p_{\text{guess}}(X|Y)_P\,. \tag{3.19}$$

The min-entropy belongs to a class of Rényi entropies that have found widespread use in information theory, and we will explore that connection in the homework. For now let us just point out that it is always smaller than the Shannon entropy.

Example. *Consider a source with joint probability distribution as in the previous example. We already exhibited a strategy that can extract a single secret bit. This is in fact optimal (as we will see) since for this distribution it is easy to compute that $H_{\min}(X|Y) = 1$. It is also worth noting that $H(X|Y) > 1$ in this case, but Shannon entropy is not the correct measure to decide how many secret bits we can extract.*

Lemma 3.3. *For any joint pmf P_{XY}, we have $H_{\min}(X|Y) \leq H(X|Y)$.*

Proof. To see this, we use Jensen's inequality on the convex function $t \mapsto -\log t$ to find

$$H_{\min}(X|Y) = -\log\left(\sum_{y \in \mathcal{Y}} P_Y(y) \max_{x \in \mathcal{X}} P_{X|Y}(x|y)\right) \tag{3.20}$$

$$\leq \sum_{y \in \mathcal{Y}} P_Y(y) \min_{x \in \mathcal{X}} \Big(-\log P_{X|Y}(x|y)\Big) \tag{3.21}$$

$$\leq \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{x \in \mathcal{X}} P_{X|Y}(x|y) \Big(-\log P_{X|Y}(x|y)\Big) = H(X|Y)\,, \tag{3.22}$$

where for the second inequality we used the fact that the minimum over x is smaller than the expectation over x under the distribution $P_{X|Y}(x|y)$. □

We will state our results using a variation of the min-entropy, the *smooth min-entropy*, which is the maximum of the min-entropy over a set of distributions that are close to P_{XY} in total variation distance.

Let P_{XY} a joint pmf and $\epsilon \in [0,1)$. We define the ϵ-smooth min-entropy of X conditioned on Y as

$$H_{\min}^{\epsilon}(X|Y)_P := \max_{\widetilde{P}_{X|Y}:\ \delta_{\text{tvd}}(\widetilde{P}_{XY},P_{XY})\leq\epsilon} H_{\min}(X|Y)_{\widetilde{P}}\,. \tag{3.23}$$

where $\widetilde{P}_{XY}(x,y) = P_Y(y)\widetilde{P}_{X|Y}(x|y)$.

We can relate the smooth entropy to the Shannon entropy again if we consider a memoryless source $(\boldsymbol{X},\boldsymbol{Y})$ producing sequences X^n and Y^n. In that case we have the following relation, which we will not prove here:

$$\forall \epsilon \in (0,1): \quad \lim_{n\to\infty} \frac{1}{n} H_{\min}^{\epsilon}(X^n|Y^n) = H(X|Y)\,. \tag{3.24}$$

3.2 *Achievability via two-universal hash functions*

There are several ways to construct extractors, including using the property of typical sets that all its elements are approximately equally likely. Here we follow a different approach (which is quite standard in cryptography) and use a random construction based on hash functions. In particular, we consider a family of two-universal hash functions $\{f_s\}_{s\in\mathcal{S}}$ where $f_s : \mathcal{X} \to \{0,1\}^L$. They are parametrised by a seed s and have the property that

$$\sum_{s\in\mathcal{S}} \frac{1}{|\mathcal{S}|} \mathbf{1}\{f_s(x) = f_s(x')\} \leq \frac{1}{2^L} \qquad \forall\, x \neq x'\,. \tag{3.25}$$

This condition can equivalently be expressed as

$$P\big[f_S(x) = f_S(x')\big] \leq \frac{1}{2^L} \qquad \forall\, x \neq x'\,. \tag{3.26}$$

where S is distributed uniformly over $\mathcal{S}$. This is the behaviour we expect from a function that produces completely random output when we take a uniformly random seed s. Such families of hash functions exist if we choose $\mathcal{S}$ large enough, but constructing them is out of the scope of this lecture.

Can you nonetheless come up with such a family for the case where $X, Z \in \{0,1\}^L$ as well? Your knowledge of finite fields might be helpful!

Let us know apply a function f_S from a two-universal family of hash functions for $S \in \mathcal{S}$ chosen uniformly at random to get an output $Z = f_S(X)$.

Theorem 3.4. *Let $\{f_s\}_s$ be a two-universal family of hash functions. Using the notation introduced above, we have*

$$\mathbb{E}\left[\delta_{\mathrm{tvd}}(P^S_{ZY}, U_Z \times P_Y)\right] \leq \frac{1}{2}\sqrt{2^{L-H_{\min}(X|Y)}}, \tag{3.27}$$

where $P^s_{ZY}(z,y) = \sum_{x\in\mathcal{X}:f_s(x)=z} P_{XY}(x,y)$ and the expectation is taken over a uniformly distributed seed S in $\mathcal{S}$.

Proof. Without loss of generality we can assume that the marginal P_Y has full support as otherwise we can just remove unused symbols.

Let us now first rewrite the left-hand side as

$$\mathbb{E}\left[\delta_{\mathrm{tvd}}(P^S_{ZY}, U_Z \times P_Y)\right] = \frac{1}{2}\mathbb{E}\left[\left\|P^S_{Z|Y}(\cdot|Y) - U_Z\right\|_1\right] \tag{3.28}$$

where the expectation on the right-hand side is over both S and Y.

We can now use the Cauchy-Schwarz inequality in Lemma 0.9 and the fact that $|Z| = 2^L$ and $U_Z(z) = 2^{-L}$ is the uniform distribution on $\mathcal{Z}$ to bound the terms

$$\left\|P^s_{Z|Y}(\cdot|y) - U_Z\right\|_1 \leq \sqrt{2^L}\left\|P^s_{Z|Y}(\cdot|y) - U_Z\right\|_2 \tag{3.29}$$

$$= \sqrt{2^L \sum_{z\in\mathcal{Z}}\left(P^s_{Z|Y}(z|y) - 2^{-L}\right)^2} \tag{3.30}$$

$$= \sqrt{2^L \sum_{z\in\mathcal{Z}} P^s_{Z|Y}(z|y)^2 - 1}\,. \tag{3.31}$$

We can now rewrite $g(y,s) = \sum_z P^s_{Z|Y}(z|y)^2$ as follows:

$$g(y,s) = \sum_z \sum_{x,x'} \mathbf{1}\{f_s(x) = z\}\mathbf{1}\{f_s(x') = z\}P_{X|Y}(x|y)P_{X|Y}(x'|y) \tag{3.32}$$

$$= \sum_{x,x'} \mathbf{1}\{f_s(x) = f_s(x')\}P_{X|Y}(x|y)P_{X|Y}(x'|y) \tag{3.33}$$

Using Jensen's inequality, the expectation over the seed S and the side information Y of the above quantity can then be taken into the square root, i.e.,

$$\mathbb{E}\left[\left\|P^s_{Z|Y=y} - U_Z\right\|_1\right] \leq \sqrt{2^L\,\mathbb{E}[g(Y,S)] - 1} \tag{3.34}$$

It remains to analyse the expectation value of g. We treat the cases where $x \neq x'$ and where $x = x'$ distinctly. In the first case we can apply the property of two-universal hash functions in (3.25). This

yields the following bound:

$$\mathbb{E}[g(Y,S)] = \sum_{x,x'} \mathbb{E}[P_{X|Y}(x|Y)P_{X|Y}(x'|Y)]\mathbb{E}[\mathbf{1}\{f_S(x) = f_S(x')\}] \quad (3.35)$$

$$\leq 2^{-L} \sum_{x \neq x'} \mathbb{E}[P_{X|Y}(x|Y)P_{X|Y}(x'|Y)] + \sum_{x} \mathbb{E}[P_{X|Y}(x|Y)P_{X|Y}(x|Y)] \quad (3.36)$$

$$\leq 2^{-L} + \sum_{y} P_Y(y) \max_x P_{X|Y}(x|y) \quad (3.37)$$

$$= 2^{-L} + p_{\text{guess}}(X|Y)\,. \quad (3.38)$$

Here, to get the last inequality we simply completed the sum to all x, x' in the first term and then used that $\sum_x P_{X|Y}(x|y) = 1$. Similarly, for the second term, we bounded one of the $P_{X|Y}(x|Y)$ with $\max_x P_{X|Y}(x|Y)$ so that the sum can be computed. Finally, plugging this into Eq. (3.34), we arrive at the desired bound. □

We can leverage this to arrive at the following result.

Theorem 3.5. *Consider a source with pmf P_{XY} and let $\epsilon \in (0,1)$. If*

$$L \leq H_{\min}^{\epsilon-\delta}(X|Y) - 2\log\frac{1}{2\delta} \quad (3.39)$$

for any $\delta \in (0,\epsilon)$, then there exists an $(\epsilon, 2^L)$-extractor for P_{XY}. This implies, in particular, that

$$L_\epsilon^*(X|Y)_P \geq \sup_{\delta\in(0,\epsilon)} H_{\min}^{\epsilon-\delta}(X|Y) - 2\log\frac{1}{\delta} + 1\,. \quad (3.40)$$

Proof. Let $\widetilde{P}_{X|Y}$ denote the distribution that achieves the maximum for the smooth min-entropy, i.e. $H_{\min}^{\epsilon-\delta}(X|Y)_P = H_{\min}(X|Y)_{\widetilde{P}}$. Theorem 3.4 applied for the source $\widetilde{P}_{XY}$ with the above choice of L yields

$$\mathbb{E}\left[\delta_{\text{tvd}}(\widetilde{P}_{ZY}^s, U_Z \times P_Y)\right] \leq \delta\,. \quad (3.41)$$

Hence, there is at least one seed value s for which this bound holds, and it remains to show that f_s constitutes an $(\epsilon, 2^L)$-extractor. However, $\delta_{\text{tvd}}(\tilde{P}_{XY}, P_{XY}) \leq \epsilon - \delta$ implies $\delta_{\text{tvd}}(\tilde{P}_{ZY}^s, P_{ZY}^s) \leq \epsilon - \delta$ by the DPI. And hence, using the triangle inequality we have

$$\delta_{\text{tvd}}(P_{ZY}^s, U_Z \times P_Y) \leq \delta_{\text{tvd}}(\widetilde{P}_{ZY}^s, P_{ZY}^s) + \delta_{\text{tvd}}(\widetilde{P}_{ZY}^s, U_Z \times P_Y) \leq \epsilon\,. \quad (3.42)$$

□

3.3 *Converse via an entropy inequality*

The converse relies on a generalization of the following lemma, which states that applying a function to a random variable cannot

increase the uncertainty about it.

Lemma 3.6. *Let $f : \mathcal{X} \to \mathcal{Z}$ be a function. Then $H(X) \geq H(f(X))$ and $H_{\min}(X) \geq H_{\min}(f(X))$.*

Proof. Let $Z = f(X)$. The joint distribution $P_{XZ}(x,z) = P_X(x)\,1\{f(x) = z\}$ satisfies

$$H(XZ) = \sum_{x,z} P_{XZ}(x,z) \log \frac{1}{P_{XZ}(x,z)} \tag{3.43}$$

$$= \sum_{x} P_X(x) \log \frac{1}{P_X(x)} = H(X)\,. \tag{3.44}$$

Hence, we conclude that $H(X) = H(XZ) = H(Z) + H(X|Z) \geq H(Z)$.

The proof for the min-entropy cannot rely on the chain rule but by inspecting the definition of the respective guessing probabilities,

$$p_{\text{guess}}(X) = \max_x P_X(x) \qquad \text{and} \tag{3.45}$$

$$p_{\text{guess}}(Z) = \max_z P_Z(z) = \max_z \sum_{x:f(x)=z} P_X(x), \tag{3.46}$$

we see that the second term is always at least as large as the first one, i.e. $p_{\text{guess}}(Z) \geq p_{\text{guess}}(X)$. This coincides with out intuition that the input of a function is at least as hard to guess as its output, since once you guessed the input you can get the output by just applying the function. The relation for the min-entropy then follows. □

It is really important that in the statement we only allow for deterministic functions, as otherwise the equality in Eq. (3.44) does not hold.

Give an example where the inequality is violated by a probabilistic function.

For our argument we need something similar to the above lemma, but for smooth min-entropy and with side information. Here again we can intuitively argue that it is at least as difficult to guess the input of a function as it is to guess the output (with equality if the function is injective). Formally, we can show the following:

Lemma 3.7. *Let $\epsilon \in [0,1)$ and $f : \mathcal{X} \to \mathcal{Z}$ be a function. Then,*

$$H^{\epsilon}_{\min}(X|Y) \geq H^{\epsilon}_{\min}(f(X)|Y)\,. \tag{3.47}$$

In the proof we will make the assumption that f is surjective. This can be avoided, but since it is not really restrictive we made it here to allow for a streamlined presentation.

Proof. The function f can be interpreted as a channel, $W_{Z|X}(z|x) = \delta_{z,f(x)}$. We can define an inverse channel, $\widetilde{W}_{X|ZY}$, that recovers the

distribution P_{XY} by Bayes' rule:

$$\widetilde{W}_{X|ZY}(x|z,y) = \frac{P_{XZ|Y}(x,z|y)}{P_{Z|Y}(z|y)} = \frac{\delta_{z,f(x)} P_{X|Y}(x|y)}{\sum_{x':f(x')=z} P_{X|Y}(x'|y)} \tag{3.48}$$

Since this channel only maps z to values of x with $f(x) = z$ it is in fact a proper right-inverse of $W_{Z|X}$ in the following sense. For any pdf Q_{ZY} on the output we define $\widetilde{Q}_{ZY}$ as the distribution resulting from first applying $\widetilde{W}_{X|YZ}$ and then $W_{Z|X}$ to Q_{ZY}. We then find that for all z, y,

Can you see what goes wrong here if the function is not surjective?

$$\widetilde{Q}_{ZY}(z,y) = \sum_{x'} W_{Z|X}(z|x') \sum_{z'} \widetilde{W}_{X|ZY}(x'|z',y) Q_{ZY}(z',y) \tag{3.49}$$

$$= \frac{\sum_{x',z'} \delta_{z,f(x')} \delta_{z',f(x')} P_{X|Y}(x'|y) Q_{ZY}(z',y)}{\sum_{x':f(x')=z} P_{X|Y}(x'|y)} = Q_{ZY}(z,y)\,. \tag{3.50}$$

Now let us assume that the distribution Q_{ZY} is optimal for the smooth min-entropy $H^\epsilon_{\min}(Z|Y)_P$, i.e. $H^\epsilon_{\min}(Z|Y)_P = H_{\min}(Z|Y)_Q$. We can then construct

$$Q_{XY}(x,y) = \sum_{z'} \widetilde{W}_{X|YZ}(x|z',y) Q_{ZY}(z',y)\,. \tag{3.51}$$

Note now that due to Eq. (3.50) the pdf $Q_{ZY}(z,y)$ is recovered by applying the function f on the register X. By the DPI for the TVD we have $\delta_{\text{tvd}}(Q_{XY}, P_{XY}) \leq \delta_{\text{tvd}}(Q_{ZY}, P_{ZY}) \leq \epsilon$. Hence,

$$H^\epsilon_{\min}(X|Y)_P \geq H_{\min}(X|Y)_Q = -\log p_{\text{guess}}(X|Y)_Q\,. \tag{3.52}$$

Now we simply use Lemma 3.6 to show that

$$p_{\text{guess}}(X|Y)_Q = \sum_y Q_Y(y)\, p_{\text{guess}}(X)_{Q^y} \tag{3.53}$$

$$\leq \sum_y Q_Y(y)\, p_{\text{guess}}(Z)_{Q^y} = p_{\text{guess}}(Z|Y)_Q\,, \tag{3.54}$$

where $Q^y_X(x) = Q_{X|Y}(x|y)$ and $Q^y_Z(z) = Q_{Z|Y}(z|y)$, respectively. Combining this with Eq. (3.52) yields the desired bound:

$$H^\epsilon_{\min}(X|Y)_P \geq H_{\min}(Z|Y)_Q = H^\epsilon_{\min}(Z|Y)_P\,. \tag{3.55}$$

□

Now we are ready to provide an upper bound on the amount of randomness that can be extracted from a source. It matches the lower bound that we derived using two-universal hash functions, and thus we know that this construction was essentially optimal.[1]

[1] To be more precise, by essentially optimal we meant that if we apply both the achievability and converse bounds to a DMS which produces independent samples from the distribution P_{XY}, then our two bounds from Theorems 3.5 and 3.8 asymptotically coincide, i.e. we can use (3.24) to establish that

$$\lim_{n\to\infty} \frac{1}{n} L^*_\epsilon(X^n|Y^n)_P = H(X|Y)\,.$$

Theorem 3.8. *Consider $\epsilon \in (0,1)$ and a source with pmf P_{XY}. Then, any $(\epsilon, 2^L)$-extractor for P_{XY} must satisfy*

$$L \leq H_{\min}^{\epsilon}(X|Y)_P \tag{3.56}$$

Or, in other words, we have $L_{\epsilon}^{*}(X|Y)_P \leq H_{\min}^{\epsilon}(X|Y)_P$.

Proof. Let us assume there exists a function f that constitutes an $(\epsilon, 2^L)$-extractor. We then necessarily have

$$\delta_{\text{tvd}}(P_{ZY}, U_Z \times P_Y) \leq \epsilon \tag{3.57}$$

for $Z = f(X)$, and thus

$$H_{\min}^{\epsilon}(Z|Y)_P \geq H_{\min}(Z|Y)_{U\times P} = H_{\min}(Z)_U = L\,, \tag{3.58}$$

where we simply evaluated the min-entropy for the distribution $U_Z \times P_Y$, which is ϵ-close to the distribution P_{ZY}. Combining this with Lemma 3.7 yields the bound $H_{\min}^{\epsilon}(X|Y)_P \geq L$, and since this holds for any $(\epsilon, 2^L)$-extractor we have shown the desired statement. □