Smooth Rényi Entropy and Applications

Renato Renner [*]   Stefan Wolf [†]

Abstract

We introduce a new entropy measure, called *smooth Rényi entropy*. The measure characterizes fundamental properties of a random variable Z, such as the amount of uniform randomness that can be extracted from Z or the minimum length of an encoding of Z. In particular, for the important case of many independent realizations of the same random experiment, smooth entropy converges to Shannon entropy.

1 Introduction

In information theory, one often wants to quantify information—usually characterized by a random variable Z with distribution P_Z—with respect to some given task. For instance, one is interested in the minimum length $|C|$ of an encoding $C = E(Z)$ of Z, such that Z can be retrieved from C (with high probability). Another such quantity is the amount $|U|$ of (almost) uniform randomness $U = F(Z)$ that can be extracted from Z by application of an extractor [9] (e.g., a random function) F.

Since it is generally hard to determine these quantities exactly, one is interested in bounds for them in terms of easily computable measures. An example for such a measure is the Shannon entropy $H(Z)$ of Z. Indeed, if Z consists of many independent realizations of the same random experiment, $H(Z)$ is asymptotically equal to the encoding length $|C|$ as well as the amount $|U|$ of extractable uniform randomness. However, for a general probability distribution P_Z, it is possible that $|U| \ll H(Z) \ll |C|$.

In this paper, we introduce a new class of entropy measures H_α^ε, called *smooth Rényi entropy*, which are obtained by a slight extension of the definition of Rényi entropy [10]. More precisely, for any parameter $\alpha \in \mathbb{R}^+$, H_α^ε is defined as a smoothed version of the Rényi entropy H_α, which means that it has the following (natural) property: For two random variables Z and Z' having almost the same probability distribution (e.g., $Z = Z'$ holds with high probability), the difference between $H_\alpha^\varepsilon(Z)$ and $H_\alpha^\varepsilon(Z')$ is small. This is generally not the case for the Rényi entropy, where, for instance, $H_\alpha(Z)$ might be much smaller than $H_\alpha(Z')$.

Another property of our new measure is that the entropy functions H_α^ε, for different values of $\alpha > 1$, only differ by some (small) additive constant. The same holds for the measures H_α^ε with $\alpha < 1$. One interesting implication of this property is that, for any random variable Z with Rényi entropy $H_\alpha(Z)$, for $\alpha > 1$, there is another random variable Z' which equals Z except with small probability and whose min-entropy $H_\infty(Z')$ is almost as large as $H_\alpha(Z)$.

As an application, we show that $H_\alpha^\varepsilon(Z)$, for $\alpha = \infty$, is a good measure for the amount of extractable uniform randomness $U = F(Z)$ (Section 7.1). In particular, together with the above-mentioned property of H_α^ε, this implies that any extractor for generating τ uniform random bits from a random variable Z with min-entropy $H_\infty(Z) = \kappa$ can also be applied to any random variable Z' with Rényi entropy $H_\alpha(Z')$ at least (roughly) κ. For $\alpha = 2$, this improves the trivial bound resulting from $H_\infty(Z') \geq \frac{1}{2}H_2(Z')$ by a factor of 2.

[*]Computer Science Department, ETH Zürich, Switzerland. Email: renner@inf.ethz.ch.

[†]Département d'Informatique et R.O., Université de Montréal, QC, Canada. Email: wolf@iro.umontreal.ca.

Similarly, $H_\alpha^\varepsilon(Z)$, for $\alpha = 0$, quantifies the minimum amount of information needed to correctly guess the value of Z (Section 7.2), which also corresponds to the minimum length of an encoding $C = E(Z)$ of Z.

2 Preliminaries

The statistical distance between two probability distributions P and Q with alphabet $\mathcal{Z}$ is defined as

$$\delta(P, Q) := \frac{1}{2} \sum_{z \in \mathcal{Z}} |P(z) - Q(z)| .$$

For two random variables Z and Z' with range $\mathcal{Z}$ distributed according to P_Z and $P_{Z'}$, respectively, we also write $\delta(Z, Z')$ instead of $\delta(P_Z, P_{Z'})$. The statistical distance satisfies the triangle inequality. Moreover, $d(Z, Z')$ cannot increase when applying a function f to its arguments, i.e.,

$$\delta(Z, Z') \geq \delta(f(Z), f(Z')) . \tag{1}$$

Note that equality holds if P_Z and $P_{Z'}$ only differ for two values $z_1, z_2 \in \mathcal{Z}$ with $f(z_1) \neq f(z_2)$. Using this, it is easy to verify that, for random variables Z, W', and any surjective function f from $\mathcal{Z}$ to $\mathcal{W}$, there exists a random variable Z' with $f(Z') = W'$ such that equality holds in (1). Hence, for any $\varepsilon \in \mathbb{R}^+ := \{r \in \mathbb{R} : r \geq 0\}$,

$$\{P_{f(Z')} : \delta(Z, Z') \leq \varepsilon\} = \{P_{W'} : \delta(f(Z), W') \leq \varepsilon\} . \tag{2}$$

The statistical distance between two pairs of random variables (Z, W) and $(Z'W')$ is bounded by

$$\max(\delta(Z, Z'), \delta(W, W')) \leq \delta(ZW, Z'W') \leq \delta(Z, Z') + \delta(W, W') . \tag{3}$$

The *non-uniformity* of a probability distribution P with range $\mathcal{Z}$ is defined by $d(P) := \delta(P, U)$ where U is the uniform probability distribution over $\mathcal{Z}$. For a random variable Z, we set $d(Z) := d(P_Z)$. For a pair of random variables Z and W, the *expected non-uniformity* of Z given W is

$$d(Z|W) := E_W[d(P_{Z|W=w})] = \sum_{w \in \mathcal{W}} P_W(w) d(P_{Z|W=w}) .$$

A straightforward calculation shows that $d(Z|W) \geq d(Z)$ and, if W is uniformly distributed, $d(ZW) = d(Z|W)$.

For a probability distribution P with range $\mathcal{Z}$ and $\alpha \in \mathbb{R}^+ \backslash \{0, 1\}$ the *Rényi entropy* (of *order* α) of P is defined as[1]

$$H_\alpha(P) := \frac{1}{1-\alpha} \log(p_\alpha(P))$$

where $p_\alpha(P) := \sum_{z \in \mathcal{Z}} P(z)^\alpha$. The definition of H_α can be extended to all parameters $\alpha \in \mathbb{R}^\infty := \mathbb{R}^+ \cup \{\infty\}$ by taking the limits $H_\alpha(P) := \lim_{\beta \to \alpha} H_\beta(P)$. This is equivalent to $H_0(P) = \log(|\{z \in \mathcal{Z} : P(z) > 0\}|)$, $H_1(P) = H(P)$, and $H_\infty(P) = -\log(p_{\max}(P))$, where $H(P)$ denotes the Shannon entropy of the probability distribution P and where $p_{\max}(P) := \max_{z \in \mathcal{Z}}(P(z))$. The Rényi entropy H_∞ is also called *min-entropy*. For a random variable Z distributed according to P_Z, we often write $H_\alpha(Z)$ instead of $H_\alpha(P_Z)$.

A well known property of the Rényi entropy is that, for $\alpha, \beta \in \mathbb{R}^\infty$,

$$\alpha \leq \beta \quad \Longleftrightarrow \quad H_\alpha(Z) + \frac{\alpha}{\alpha - 1} \log(N) \geq H_\beta(Z) + \frac{\beta}{\beta - 1} \log(N) . \tag{4}$$

[1] All logarithms in this paper are binary.

where $N = \sum_z P_Z(z)$ is a normalization constant. Moreover, the Rényi entropy is subadditive, i.e., for random variables Z and W,

$$H_\alpha(ZW) \leq H_\alpha(Z) + H_\alpha(W) \ , \tag{5}$$

where equality in (5) holds if Z and W are independent.

3 Smooth Rényi Entropy: Definitions

Let $\mathcal{P}(\mathcal{X})$ be the set of non-negative real functions on $\mathcal{X}$.

3.1 Definition by events

Let X be a random variable on $\mathcal{X}$ with probability distribution P_X and let $\mathcal{E}$ be an event defined by $P_{\mathcal{E}|X=x}$, for $x \in \mathcal{X}$. We denote by $P_{X\mathcal{E}}$ the generalized (?) probability distribution defined by $P_{X\mathcal{E}}(x) := \mathrm{Prob}[(X = x) \wedge \mathcal{E}] = P_X(x) \cdot P_{\mathcal{E}|X=x}$.

For $\alpha \in (0,1) \cup (1,\infty)$ and $p \in \mathcal{P}(\mathcal{X})$, let

$$r_\alpha(p) := \sum_{x \in \mathcal{X}} p(x)^\alpha \ .$$

For a random variable X and $\varepsilon \geq 0$, define

$$r_\alpha^\varepsilon(X) := \inf_{\mathcal{E}: P_\mathcal{E} \geq 1-\varepsilon} r_\alpha(P_{X\mathcal{E}}) \ .$$

For $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, let

$$r_\alpha^\varepsilon(X|Y) = \inf_{\mathcal{E}: P_\mathcal{E} \geq 1-\varepsilon} \max_{y \in \mathcal{Y}} \bigl(r_\alpha(P_{X\mathcal{E}|Y=y}\bigr) \ ,$$

where $P_{X\mathcal{E}|Y=y}$ is defined by $P_{X\mathcal{E}|Y=y}(x) := \mathrm{Prob}[(X = x) \wedge \mathcal{E}|Y = y]$.

3.2 Definition by variational distance

For any $\varepsilon \geq 0$ and $p \in \mathcal{P}_\mathcal{X}$, let $\mathcal{B}^\varepsilon(p) := \{q \in \mathcal{P}_\mathcal{X} : \delta(p,q) \leq \varepsilon\}$.

Let $\alpha \in (0,1) \cup (1,\infty)$. For $p \in \mathcal{P}_\mathcal{X}$, let

$$r_\alpha(p) := \sum_{x \in \mathcal{X}} p(x)^\alpha \ .$$

Let $P_X \in \mathcal{P}(\mathcal{X})$. For $\varepsilon \geq 0$, define

$$r_\alpha^\varepsilon(p) := \inf_{p' \in \mathcal{B}^\varepsilon(p)} r_\alpha(p') \ .$$

Adapting the notation commonly used for entropies, we will also write $r_\alpha^\varepsilon(X)$ instead of $r_\alpha^\varepsilon(P_X)$, and, more generally, for an event $\mathcal{E}$, $r_\alpha^\varepsilon(X|\mathcal{E})$ instead of $r_\alpha^\varepsilon(P_{X|\mathcal{E}})$. For $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, let

$$r_\alpha^\varepsilon(X|Y) = \inf_{(\varepsilon_y):\, \mathbf{E}[\varepsilon_y]=\varepsilon} \max_{y \in \mathcal{Y}} \bigl(r_\alpha^{\varepsilon_y}(X|Y = y)\bigr) \ ,$$

where the infimum ranges over all tuples $(\varepsilon_y)_{y \in \mathcal{Y}}$ of non-negative real numbers $\varepsilon_y \in \mathbb{R}^+$ such that $\mathbf{E}_{y \leftarrow P_Y}[\varepsilon_y] = \varepsilon$.

3.3 Equivalence of the two definitions

The two definitions above are equivalent.

3.4 General part of the definition

Definition 3.1. Let $\varepsilon \geq 0$. The *ε-smooth Rényi entropy* (of *order* α) of a random variable X is defined as

$$H_\alpha^\varepsilon(X) := \frac{1}{1-\alpha} \log r_\alpha^\varepsilon(X) \ .$$

More generally, the *ε-smooth Rényi entropy* (of *order* α) of X *conditioned* on Y is defined as

$$H_\alpha^\varepsilon(X|Y) := \frac{1}{1-\alpha} \log r_\alpha^\varepsilon(X|Y) \ .$$

For $\alpha \in \{0, \infty\}$, the ε-smooth Rényi entropy is defined by a limes, i.e.,

$$H_0^\varepsilon(X|Y) := \lim_{\alpha \to 0} H_\alpha^\varepsilon(X|Y) \tag{6}$$

$$H_\infty^\varepsilon(X|Y) := \lim_{\alpha \to \infty} H_\alpha^\varepsilon(X|Y) \ . \tag{7}$$

4 Basic properties of smooth Rényi entropy

4.1 Relation between smooth Rényi entropies of different orders

The following is a generalization of the implication

$$\varepsilon \overset{\alpha}{\leq} \varepsilon' \quad \Longleftrightarrow \quad H_\alpha^\varepsilon(Z) \leq H_\alpha^{\varepsilon'}(Z) \ .$$

Lemma 4.1. *For any random variable Z and $\varepsilon, \varepsilon' \in \mathbb{R}^+$, for $\alpha \leq \beta < 1$ or $1 < \alpha \leq \beta$,*

$$H_\alpha^\varepsilon(Z) + \frac{\alpha}{\alpha-1} \log(1-\varepsilon) \geq H_\beta^\varepsilon(Z) + \frac{\beta}{\beta-1} \log(1-\varepsilon)$$

and, for $\alpha < 1 < \beta$,

$$H_\alpha^\varepsilon(Z) + \frac{\alpha}{\alpha-1} \log(1-\varepsilon-\varepsilon') \geq H_\beta^{\varepsilon'}(Z) + \frac{\beta}{\beta-1} \log(1-\varepsilon-\varepsilon')$$

Proof. Assume that $\alpha \leq \beta < 1$. Then there exists a non-negative real function P such that

$$H_\alpha(P) \leq H_\alpha^\varepsilon(P_Z)$$

and $\sum_z P(z) \geq 1 - \varepsilon$. Hence, using (4),

$$\begin{aligned} H_\alpha^\varepsilon(P_Z) + \frac{\alpha}{\alpha-1} \log(1-\varepsilon) &\geq H_\alpha(P) + \frac{\alpha}{\alpha-1} \log(1-\varepsilon) \\ &\geq H_\beta(P) + \frac{\beta}{\beta-1} \log(1-\varepsilon) \\ &\geq H_\beta^{\varepsilon'}(P_Z) + \frac{\beta}{\beta-1} \log(1-\varepsilon) \end{aligned}$$

The proof for $1 < \alpha \leq \beta$ follows similar lines.

Finally, for $\alpha < 1 < \beta$, observe that there exists a non-negative real function P such that

$$H_\alpha(P) \leq H_\alpha^\varepsilon(P_Z)$$
$$H_\beta(P) \geq H_\beta^{\varepsilon'}(P_Z)$$

and $\sum_z P(z) \geq 1 - \varepsilon - \varepsilon'$. Hence, using (4)

$$\begin{aligned} H_\alpha^\varepsilon(P_Z) + \frac{\alpha}{\alpha-1} \log(1-\varepsilon-\varepsilon') &\geq H_\alpha(P) + \frac{\alpha}{\alpha-1} \log(1-\varepsilon-\varepsilon') \\ &\geq H_\beta(P) + \frac{\beta}{\beta-1} \log(1-\varepsilon-\varepsilon') \\ &\geq H_\beta^{\varepsilon'}(P_Z) + \frac{\beta}{\beta-1} \log(1-\varepsilon-\varepsilon') \ . \end{aligned}$$

$\square$

4.2 Two classes of smooth Rényi entropies: $\alpha < 1$ and $\alpha > 1$

The following lemma, together with Lemma 4.1, implies that the values of the smooth Rényi entropy of order α of a random variable Z, for different values of the parameter $\alpha > 1$, cannot differ by more than a small additive constant (independent of P_Z).

Lemma 4.2. *For any random variable Z, $\alpha > 1$, and $\varepsilon > 0$,*

$$H_\infty^\varepsilon(Z) \geq H_\alpha(Z) + \frac{\log(\varepsilon)}{\alpha - 1}$$

Proof. Let $P := P_Z$ and let p be a fixed probability in the interval $[1/|\mathcal{Z}|, p_{\max}(P)]$. It is easy to see that there is a probability distribution Q_p such that $p_{\max}(Q_p) = p$ and

$$\delta(P, Q_p) = \sum_{z \in \mathcal{Z}_p} P(z) - p \ , \tag{8}$$

where $\mathcal{Z}_p = \{z \in \mathcal{Z} : P(z) \geq p\}$. (E.g., Q_p can be obtained from P by cutting all probabilities larger than p while increasing the values of the smallest probabilities.)

By the definition of $\mathcal{Z}_p$, we obtain

$$p_\alpha(P) = \sum_{z \in \mathcal{Z}} P(z)^\alpha \geq \sum_{z \in \mathcal{Z}_p} P(z)^\alpha \geq p^{\alpha-1} \sum_{z \in \mathcal{Z}_p} P(x) \geq p^{\alpha-1} \cdot \delta(P, Q_p) \ .$$

where the last inequality is a direct consequence of (8). Hence,

$$-\log p_\alpha(P) \leq -\log \delta(P, Q_p) - (\alpha - 1) \log p$$

and, because of $p_{\max}(Q_p) = p$,

$$H_\alpha(P) \leq -\frac{1}{\alpha - 1} \log \delta(P, Q_p) + H_\infty(Q_p) \ .$$

The assertion follows by choosing p such that $\delta(P, Q_p) = \varepsilon$. □

In particular, for any random variable Z with Rényi entropy $H_\alpha(Z)$, for $\alpha > 1$, there is another random variable Z' with (roughly) the same amount of min-entropy $H_\infty(Z')$, which has almost the same probability distribution as Z.

This has important implications for privacy amplification (cf. [1] for definitions and references) which is used in information-theoretic secret-key agreement to transform a weakly secure key $\tilde{S}$ into a highly secure key S. It is well known that the Rényi entropy of order 2 of $\tilde{S}$ is a lower bound for the number $|S|$ of highly secure key bits which can be generated from $|S|$ by applying two-universal hashing. It follows directly from Lemma 4.2 that the same amount of secret bits can be obtained by applying any optimal strong extractor (cf. Definition 7.1 below) to $\tilde{S}$.

The following result is in a certain sense the converse of Lemma 4.2 for the case $\alpha < 1$. (The proofs might be combined.)

Lemma 4.3. *For any random variable Z, $\alpha < 1$, and $\varepsilon > 0$,*

$$H_0^\varepsilon(Z) \leq H_\alpha(Z) + \frac{\log(\varepsilon)}{\alpha - 1} \ .$$

Proof. Let $P := P_Z$ and let p be a fixed probability in the interval $[1/|\mathcal{Z}|, p_{\max}(P)]$. It is easy to see that there is a probability distribution Q_p such that $p_{\max}(Q_p) = p$ and

$$\delta(P, Q_p) = \sum_{z \in \mathcal{Z}_p} P(z) - p \ , \tag{9}$$

where $\mathcal{Z}_p = \{z \in \mathcal{Z} : P(z) \geq p\}$. (E.g., Q_p can be obtained from P by cutting all probabilities larger than p while increasing the values of the smallest probabilities.)

By the definition of $\mathcal{Z}_p$, we obtain

$$p_\alpha(P) = \sum_{z \in \mathcal{Z}} P(z)^\alpha \geq \sum_{z \in \mathcal{Z}_p} P(z)^\alpha \geq p^{\alpha-1} \sum_{z \in \mathcal{Z}_p} P(x) \geq p^{\alpha-1} \cdot \delta(P, Q_p) \ .$$

where the last inequality is a direct consequence of (9). Hence,

$$-\log p_\alpha(P) \leq -\log \delta(P, Q_p) - (\alpha - 1) \log p$$

and, because of $p_{\max}(Q_p) = p$,

$$H_\alpha(P) \leq -\frac{1}{\alpha - 1} \log \delta(P, Q_p) + H_\infty(Q_p) \ .$$

The assertion follows by choosing p such that $\delta(P, Q_p) = \varepsilon$. □

5 Some technical properties

The following properties are probably not so important for the EC-submission.

5.1 Conditioning on events

Lemma 5.1. *Let X and Y be random variables, let $\varepsilon, \varepsilon' \in \mathbb{R}^+$, and $\mathcal{E}$ be an event with* $\mathrm{Prob}[\mathcal{E}] \geq 1 - \varepsilon$. *Then, for any parameter α,*

$$H_\alpha^{\varepsilon+\varepsilon'}(X|Y) \overset{\alpha}{\leq} H_\alpha^{\varepsilon'}(X|Y, \mathcal{E}) \ .$$

Proof. Let $\mathcal{E}'$ be an event with $\mathrm{Prob}[\mathcal{E}'] \geq 1 - \varepsilon$ such that

$$\max_y r_\alpha(P_{X,\mathcal{E}'|Y=y,\mathcal{E}}) \leq r_\alpha^{\varepsilon'}(X|Y, \mathcal{E}) \ .$$

Then

$$r_\alpha^{\varepsilon+\varepsilon'}(X|Y) \leq \max_y r_\alpha(P_{X,\mathcal{E},\mathcal{E}'|Y=y}) \leq \max_y r_\alpha(P_{X,\mathcal{E}'|Y=y,\mathcal{E}}) \leq r_\alpha^{\varepsilon'}(X|Y, \mathcal{E}) \ .$$

□

Lemma 5.2. *Let X and Y be random variables and let $\varepsilon \in \mathbb{R}^+$. Then, for any parameter α, $y \in \mathcal{Y}$, and $p := P_Y(y)$,*

$$H_\alpha^{\varepsilon}(X|Y = y) \overset{\alpha}{\leq} H_\alpha^{p\varepsilon}(X|Y) \ .$$

Proof. Let $\mathcal{E}$ be an event with $\mathrm{Prob}[\mathcal{E}] \geq 1 - \varepsilon p$ such that

$$\max_y r_\alpha(P_{X,\mathcal{E}|Y=y}) \leq r_\alpha^{\varepsilon p}(X|Y) \ .$$

Then, since

$$P_{\overline{\mathcal{E}}|Y=y} = \frac{P_{Y,\overline{\mathcal{E}}}(y)}{P_Y(y)} \leq \frac{P_{\overline{\mathcal{E}}}}{p} \leq \varepsilon \ ,$$

we conclude

$$r_\alpha^{\varepsilon}(X|Y = y) \leq r_\alpha(P_{X,\mathcal{E}|Y=y}) \leq \max_y r_\alpha(P_{X,\mathcal{E}|Y=y}) \leq r_\alpha^{\varepsilon p}(X|Y) \ .$$

□

5.2 Less important properties

Lemma 5.3. *Let $\varepsilon, \varepsilon' \geq 0$ and let P_{XY} and $P_{X'Y'}$ be probability distributions such that $P_{X'Y'} \in \mathcal{B}^{\varepsilon}(P_{XY})$ and $P_{Y'} \leq P_Y$. Then*[2]*,*

$$H_\alpha^{\varepsilon+\varepsilon'}(X|Y) \overset{\alpha}{\leq} H_\alpha^{\varepsilon'}(X'|Y') .$$

Proof. The assertion of the lemma is equivalent to

$$r_\alpha^{\varepsilon+\varepsilon'}(X|Y) \leq r_\alpha^{\varepsilon'}(X'|Y') .$$

Note first that, for two probability distributions P_X and $P_{X'}$ with $\delta(P_X, P_{X'}) \leq \varepsilon$,

$$r_\alpha^{\varepsilon+\varepsilon'}(X) \leq r_\alpha^{\varepsilon'}(X') . \tag{10}$$

To see this, it suffices to verify that $P_{\tilde{X}} \in \mathcal{B}^{\varepsilon}(P_{X'})$ implies $P_{\tilde{X}} \in \mathcal{B}^{\varepsilon+\varepsilon'}(P_X)$, which is a direct consequence of the triangle inequality.

For any $y \in \mathcal{Y}$, let $p_y \in \mathcal{P}(\mathcal{X})$ be defined by $p_y(x) := \min(P_{X|Y=y}(x), P_{X'|Y'=y}(x))$, and let $\varepsilon_y := \delta(P_{X|Y=y}, p_y)$. Then

$$\mathop{\mathbf{E}}_{y \leftarrow P_Y}[\varepsilon_y] = \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{x \in \mathcal{X}} \bigl|P_{X|Y=y}(x) - p_y(x)\bigr| = \sum_{\substack{(x,y) \in \mathcal{X} \times \mathcal{Y} \\ P_{X|Y=y}(x) \geq P_{X'|Y'=y}(x)}} \bigl|P_{XY}(x,y) - P_{X'|Y'=y}(x) P_Y(y)\bigr| .$$

Using the fact that $P_Y(y) \geq P_{Y'}(y)$, we have

$$P_{X'|Y'=y}(x) P_Y(y) = P_{X'Y'}(x,y) \frac{P_Y(y)}{P_{Y'}(y)} \geq P_{X'Y'}(x,y) ,$$

and hence,

$$\mathop{\mathbf{E}}_{y \leftarrow P_Y}[\varepsilon_y] \leq \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \bigl|P_{XY}(x,y) - P_{X'Y'}(x,y)\bigr| .$$

Let now (ε'_y) such that $\mathbf{E}_{y \leftarrow P_{Y'}}[\varepsilon'_y] = \varepsilon'$. Since, for any $y \in \mathcal{Y}$, $p_y \leq P_{X'|Y'=y}$, we find, using (10),

$$r_\alpha^{\varepsilon_y+\varepsilon'_y}(P_{X|Y=y}) \leq r_\alpha^{\varepsilon_y}(p_y) \leq r_\alpha^{\varepsilon_y}(P_{X'|Y'=y})$$

...

□

The following lemma is a direct consequence of Lemma **??** and the union bound.

Lemma 5.4. *Let $X_1, \ldots X_n$ be random variables, let $\alpha_1, \ldots, \alpha_n \in \mathbb{R}^+$, and let $\varepsilon_1, \ldots, \varepsilon_n \in \mathbb{R}^+$. Then there exists an event $\mathcal{E}$ with* $\mathrm{Prob}[\mathcal{E}] \geq 1 - \sum_i \varepsilon_i$ *such that, for all $i \in [n]$,*

$$r_{\alpha_i}(P_{X_i, \mathcal{E}}) \leq r_{\alpha_i}^{\varepsilon_i}(X_i) .$$

6 Relation to Shannon entropy

6.1 Asymtotic equivalence for product distributions

Interestingly, one obtains as an immediate consequence of the asymptotic equipartition property (AEP) (cf. [5]) that, for many independent realizations of a random experiment, smooth Rényi entropy is asymptotically equal to Shannon entropy.

[2] We write $a \overset{\alpha}{\leq} b$ to denote the inequality $a \leq b$, for $\alpha < 1$, and $a \geq b$, for $\alpha > 1$.

Lemma 6.1. *Let Z be a random variable with distribution P_Z and let $Z^n := (Z_1, \ldots, Z_n)$ be an n-tuple of independent random variables Z_i distributed according to P_Z. Then, for $\alpha \in \mathbb{R}^\infty$,*

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{H_\alpha^\varepsilon(Z^n)}{n} = H(Z) \ .$$

Proof. Because of Lemma 4.1, it suffices to prove that the assertion holds for $\alpha = 0$ and $\alpha = \infty$. This is however a direct consequence of the asymptotic equipartition property (AEP) [5]. □

6.2 Shannon-style properties

6.2.1 Chain rule in one direction

Lemma 6.2. *Let X and Y be random variables and let $\varepsilon, \varepsilon', \varepsilon'' \in \mathbb{R}^+$, $\varepsilon > 0$. Then, for $\alpha < 1 < \beta$,*

$$H_\alpha^{\varepsilon+\varepsilon'+\varepsilon''}(X|Y) < H_\alpha^{\varepsilon'}(XY) - H_\beta^{\varepsilon''}(Y) + \frac{\beta - \alpha}{(1-\alpha)(\beta-1)} \log(\frac{1}{\varepsilon}) \ ,$$

and, similarly, for $\alpha > 1 > \beta$,

$$H_\alpha^{\varepsilon+\varepsilon'+\varepsilon''}(X|Y) > H_\alpha^{\varepsilon'}(XY) - H_\beta^{\varepsilon''}(Y) - \frac{\alpha - \beta}{(\alpha-1)(1-\beta)} \log(\frac{1}{\varepsilon}) \ .$$

Proof. Using the definition of smooth Rényi entropy, the assertion can be rewritten as

$$\log r_\alpha^{\varepsilon+\varepsilon'+\varepsilon''}(X|Y) < \log r_\alpha^{\varepsilon'}(XY) + \frac{1-\alpha}{\beta-1} \log r_\beta^{\varepsilon''}(Y) + \frac{\beta-\alpha}{\beta-1} \log \frac{1}{\varepsilon} \ . \tag{11}$$

According to Lemma 5.4, there exists an event $\mathcal{E}$ with $\Pr[\mathcal{E}] \geq 1 - \varepsilon' - \varepsilon''$ such that

$$r_\alpha(P_{XY,\mathcal{E}}) \leq r_\alpha^{\varepsilon'}(XY)$$

and

$$r_\beta(P_{Y,\mathcal{E}}) \leq r_\beta^{\varepsilon''}(Y) \ .$$

For any $y \in \mathcal{Y}$, let $\bar{r}_y := r_\alpha(P_{X,\mathcal{E}|Y=y})$. Since (11) is independent of the labeling of the values in $\mathcal{Y}$, we can assume without loss of generality that $\mathcal{Y} = \{1, \ldots, n\}$, for $n := |\mathcal{Y}|$, and that $\bar{r}_y > \bar{r}_{y'} \implies y > y'$.

Let $\bar{y} \in \mathcal{Y}$ be the minimum value such that $\Pr[Y > \bar{y}, \mathcal{E}] \leq \varepsilon$. In particular,

$$\Pr[Y \geq \bar{y}, \mathcal{E}] = \Pr[Y > \bar{y} - 1, \mathcal{E}] > \varepsilon \ . \tag{12}$$

Let $\mathcal{E}'$ be the event that $Y \leq \bar{y}$. We then have $\text{Prob}[\overline{\mathcal{E}'}, \mathcal{E}] \leq \varepsilon$ and thus

$$\text{Prob}[\mathcal{E}', \mathcal{E}] = 1 - \text{Prob}[\overline{\mathcal{E}}] - \text{Prob}[\overline{\mathcal{E}'}, \mathcal{E}] \geq 1 - \varepsilon - \varepsilon' - \varepsilon'' \ .$$

Hence, since $r_\alpha(P_{X,\mathcal{E},\mathcal{E}'|Y=y}) = 0$, for $y > \bar{y}$,

$$r_\alpha^{\varepsilon+\varepsilon'+\varepsilon''}(X|Y) \leq \max_{y \in \mathcal{Y}} r_\alpha(P_{X,\mathcal{E},\mathcal{E}'|Y=y}) \leq \max_{y \leq \bar{y}} r_\alpha(P_{X,\mathcal{E}|Y=y}) \leq \bar{r}$$

where $\bar{r} = r_\alpha(P_{X,\mathcal{E}|Y=\bar{y}})$. It thus remains to be proven that

$$\log(\bar{r}) < \log\big(r_\alpha(P_{XY,\mathcal{E}})\big) + \frac{1-\alpha}{\beta-1} \log\big(r_\beta(P_{Y,\mathcal{E}})\big) - \frac{\beta-\alpha}{\beta-1} \log(\varepsilon) \ . \tag{13}$$

Let $s := \sum_{y=\bar{y}}^n P_Y(y)^\alpha$. Then,

$$\bar{r} \cdot s = \sum_{y=\bar{y}}^n \bar{r} P_Y(y)^\alpha \leq \sum_{y=\bar{y}}^n \bar{r}_y P_Y(y)^\alpha \leq \sum_{y=1}^n \bar{r}_y P_Y(y)^\alpha$$

where the first inequality follows from the fact that $\bar{r}_y \geq \bar{r}$ for all $y \geq \bar{y}$. Inserting the definition of $\bar{r}_y$ into this inequality, we get

$$\bar{r} \leq \frac{1}{s} \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{XY,\mathcal{E}}(x,y)^\alpha = \frac{r_\alpha(P_{XY,\mathcal{E}})}{s} ,$$

i.e.,

$$\log(\bar{r}) \leq \log\bigl(r_\alpha(P_{XY})\bigr) - \log(s) . \tag{14}$$

To find a bound on s, let $p_y := P_{Y,\mathcal{E}}(y)$, $p := \frac{\beta-\alpha}{\beta-1}$, $q := \frac{\beta-\alpha}{1-\alpha}$, and $\gamma := \frac{\alpha(\beta-1)}{\beta-\alpha}$, i.e., $\gamma p = \alpha$ and $(1-\gamma)q = \beta$. Then, since $\frac{1}{p} + \frac{1}{q} = 1$, we can apply Hölder's, inequality leading to

$$\begin{aligned}
\sqrt[p]{s} \cdot \sqrt[q]{r_\beta(P_{Y,\mathcal{E}})} &\geq \sqrt[p]{\sum_{y=\bar{y}}^{n} (p_y)^\alpha} \cdot \sqrt[q]{\sum_{y=\bar{y}}^{n} (p_y)^\beta} \\
&= \sqrt[p]{\sum_{y=\bar{y}}^{n} \bigl((p_y)^\gamma\bigr)^p} \cdot \sqrt[q]{\sum_{y=\bar{y}}^{n} \bigl((p_y)^{1-\gamma}\bigr)^q} \\
&\geq \sum_{y=\bar{y}}^{n} (p_y)^\gamma (p_y)^{1-\gamma} = \sum_{y=\bar{y}}^{n} p_y = \Pr[Y \geq \bar{y}, \mathcal{E}] > \varepsilon .
\end{aligned}$$

Hence,

$$\log(s) > p \log(\varepsilon) - \frac{p}{q} \log\bigl(r_\beta(P_{Y,\mathcal{E}})\bigr) = \frac{\beta-\alpha}{\beta-1} \log(\varepsilon) - \frac{1-\alpha}{\beta-1} \log\bigl(r_\beta(P_{Y,\mathcal{E}})\bigr) .$$

Combining this with (14) implies (13) and thus concludes the proof. □

6.2.2 Chain rule in the other (more trivial) direction

Lemma 6.3. *Let X and Y be random variables and let $\varepsilon, \varepsilon' \in \mathbb{R}^+$. Then, for any parameter α,*

$$H_\alpha^{\varepsilon+\varepsilon'}(XY) \overset{\alpha}{\lesssim} H_\alpha^\varepsilon(X|Y) + H_\alpha^{\varepsilon'}(Y) .$$

Proof. Let $\mathcal{E}$ be an event with $\mathrm{Prob}[\mathcal{E}] \geq 1-\varepsilon$ such that and

$$\max_y r_\alpha(P_{X,\mathcal{E}|Y=y}) \leq r_\alpha^\varepsilon(X|Y) .$$

Similarly, let $\mathcal{E}'$ be an event with $\Pr[\mathcal{E}'] \geq 1-\varepsilon'$ such that $\mathcal{E}' \leftarrow Y \leftarrow (X,\mathcal{E})$ and

$$r_\alpha(P_{Y,\mathcal{E}'}) \leq r_\alpha^\varepsilon(Y) .$$

Since $\mathrm{Prob}[\mathcal{E},\mathcal{E}'] \geq 1-\varepsilon-\varepsilon'$, we have $r_\alpha^{\varepsilon+\varepsilon'}(XY) \leq r_\alpha(P_{XY,\mathcal{E},\mathcal{E}'})$, i.e., it suffices to show that

$$r_\alpha(P_{XY,\mathcal{E},\mathcal{E}'}) \leq r_\alpha(P_{Y,\mathcal{E}'}) \max_y r_\alpha(P_{X,\mathcal{E}|Y=y}) .$$

The first term can be rewritten as

$$r_\alpha(P_{XY,\mathcal{E},\mathcal{E}'}) = \sum_{x,y} P_{XY,\mathcal{E},\mathcal{E}'}(x,y)^\alpha = \sum_{x,y} \left(\frac{P_{XY,\mathcal{E},\mathcal{E}'}(x,y)}{P_{Y,\mathcal{E}'}(y)}\right)^\alpha P_{Y,\mathcal{E}'}(y)^\alpha ,$$

where the last sum ranges over all values y such that $P_{Y,\mathcal{E}'}(y) > 0$. Note that, for any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ with $P_{Y,\mathcal{E}'}(y) > 0$,

$$\frac{P_{XY,\mathcal{E},\mathcal{E}'}(x,y)}{P_{Y,\mathcal{E}'}(y)} = \frac{P_{XY,\mathcal{E}}(x,y)}{P_Y(y)} = P_{X,\mathcal{E}|Y=y}(x) .$$

Hence

$$\begin{aligned}
r_\alpha(P_{XY,\mathcal{E},\mathcal{E}'}) &\leq \sum_{x,y} P_{Y,\mathcal{E}'}(y)^\alpha P_{X,\mathcal{E}|Y=y}(x)^\alpha \\
&\leq \sum_{y} P_{Y,\mathcal{E}'}(y)^\alpha \max_y \sum_x P_{X,\mathcal{E}|Y=y}(x)^\alpha \\
&= r_\alpha(P_{Y,\mathcal{E}'}) \max_y r_\alpha(P_{X,\mathcal{E}|Y=y})
\end{aligned}$$

which concludes the proof. □

Corollary 6.4. *Let X, Y, and Z be random variables and let $\varepsilon, \varepsilon' \in \mathbb{R}^+$. Then, for any parameter α,*

$$H_\alpha^{\varepsilon+\varepsilon'}(XY|Z) \overset{\alpha}{\leq} H_\alpha^{\varepsilon}(X|YZ) + H_\alpha^{\varepsilon'}(Y|Z) \ .$$

Proof. According to the definition of conditional smooth Rény entropy, there exists $\varepsilon_z, \varepsilon'_z \in \mathbb{R}^+$, for $z \in \mathcal{Z}$, such that $\mathbf{E}[\varepsilon_z] = \varepsilon$, $\mathbf{E}[\varepsilon'_z] = \varepsilon'$, and

$$\begin{aligned}
r_\alpha^\varepsilon(X|YZ) &= \max_{z \in \mathcal{Z}} r_\alpha^{\varepsilon_z}(X|Y, Z = z) \\
r_\alpha^\varepsilon(Y|Z) &= \max_{z \in \mathcal{Z}} r_\alpha^{\varepsilon'_z}(Y|Z = z) \ .
\end{aligned}$$

Note that $\mathbf{E}[\varepsilon_z + \varepsilon'_z] \leq \varepsilon + \varepsilon'$. Hence, by Lemma 6.3,

$$\begin{aligned}
\log r_\alpha^{\varepsilon+\varepsilon'}(XY|Z) &\leq \log \max_{z \in z\mathcal{Z}} r_\alpha^{\varepsilon_z+\varepsilon'_z}(XY|Z = z) \\
&\leq \log \max_{z \in z\mathcal{Z}} r_\alpha^{\varepsilon_z}(X|Y, Z = z) + \log \max_{z \in \mathcal{Z}} r_\alpha^{\varepsilon'_z}(Y|Z = z) \\
&\leq \log r_\alpha^\varepsilon(X|YZ) + \log r_\alpha^\varepsilon(Y|Z) \ ,
\end{aligned}$$

which concludes the proof. □

6.2.3 Subadditivity

Lemma 6.5. *Let X and Y be random variables and let $\varepsilon \in \mathbb{R}^+$. Then, for $\alpha < 1$,*

$$H_\alpha^{\varepsilon+\varepsilon'}(XY) \leq H_\alpha^\varepsilon(X) + H_0^{\varepsilon'}(Y) \ ,$$

and, similarly, for $\alpha > 1$,

$$H_\alpha^\varepsilon(XY) \leq H_\alpha^{\varepsilon+\varepsilon'}(X) + H_0^{\varepsilon'}(Y) \ .$$

Proof. Assume first that $\alpha < 1$. Let $\mathcal{E}$ be an event with $\text{Prob}[\mathcal{E}] \geq 1 - \varepsilon$ such that

$$r_\alpha(P_{X,\mathcal{E}}) \leq r_\alpha^\varepsilon(X)$$

and let $\mathcal{E}'$ be an event with $\text{Prob}[\mathcal{E}'] \geq 1 - \varepsilon'$ such that

$$r_0(P_{Y,\mathcal{E}'}) \leq r_0^{\varepsilon'}(Y) \ .$$

Since $r_0(P_{Y,\mathcal{E}'}) = r_0(P_{Y|\mathcal{E}'})$, this implies $H_0(Y|\mathcal{E}') \leq H_0^{\varepsilon'}(Y)$. Then

$$r_\alpha^{\varepsilon+\varepsilon'}(XY) \leq r_\alpha(P_{XY,\mathcal{E},\mathcal{E}'}) = \sum_{x,y} P_{X,\mathcal{E},\mathcal{E}'}(x)^\alpha P_{Y|X=x,\mathcal{E},\mathcal{E}'}(y)^\alpha \leq r_\alpha(P_{X,\mathcal{E},\mathcal{E}'}) \max_x r_\alpha(P_{Y|X=x,\mathcal{E},\mathcal{E}'}) \ .$$

Hence, since $r_\alpha(P_{X,\mathcal{E},\mathcal{E}'}) \leq r_\alpha(P_{X,\mathcal{E}}) \leq r_\alpha^\varepsilon(X)$, we find

$$H_\alpha^{\varepsilon+\varepsilon'}(XY) \leq H_\alpha^\varepsilon(X) + \max_x H_\alpha(Y|X = x, \mathcal{E}, \mathcal{E}') \ .$$

The assertion then follows from

$$H_\alpha(Y|X=x,\mathcal{E},\mathcal{E}') \leq H_0(Y|X=x,\mathcal{E},\mathcal{E}') \leq H_0(Y|\mathcal{E}') \leq H_0^{\varepsilon'}(Y) \ . \tag{15}$$

To prove the statement for $\alpha > 1$, let $\mathcal{E}$ be an event with $\mathrm{Prob}[\mathcal{E}] \geq 1-\varepsilon$ such that

$$r_\alpha(P_{XY,\mathcal{E}}) \leq r_\alpha^\varepsilon(XY)$$

and let $\mathcal{E}'$ be defined as above. Since

$$r_\alpha(P_{XY,\mathcal{E}}) \geq r_\alpha(P_{XY,\mathcal{E},\mathcal{E}'}) = \sum_{x,y} P_{X,\mathcal{E},\mathcal{E}'}(x)^\alpha P_{Y|X=x,\mathcal{E},\mathcal{E}'}(y)^\alpha \geq r_\alpha(P_{X,\mathcal{E},\mathcal{E}'}) \min_x r_\alpha(P_{Y|X=x,\mathcal{E},\mathcal{E}'})$$

and $r_\alpha(P_{X,\mathcal{E},\mathcal{E}'}) \geq r_\alpha^{\varepsilon+\varepsilon'}(X)$, we find

$$H_\alpha^\varepsilon(XY) \leq H_\alpha^{\varepsilon+\varepsilon'}(X) + \max_x H_\alpha(Y|X=x,\mathcal{E},\mathcal{E}') \ .$$

Using (15) then concludes the proof. □

6.2.4 The entropy of a part can not be larger then the entropy of the whole

Lemma 6.6. *Let X and Y be random variables and let $\varepsilon \in \mathbb{R}^+$. Then, for $\alpha \in (0,1) \cup (1,\infty)$,*

$$H_\alpha^\varepsilon(X) \leq H_\alpha^\varepsilon(XY) \ .$$

Proof. For $\alpha > 1$, this is an immediate consequence of Lemma 6.3 and the fact that $H_\alpha(Y|X) \geq 0$. To prove the assertion for $\alpha < 1$, let $\mathcal{E}$ be an event with $\mathrm{Prob}[\mathcal{E}] \geq 1-\varepsilon$ such that

$$r_\alpha(P_{XY,\mathcal{E}}) \leq r_\alpha^\varepsilon(XY) \ .$$

Then,

$$r_\alpha^\varepsilon(X) \leq r_\alpha(P_{X,\mathcal{E}}) \leq r_\alpha(P_{XY,\mathcal{E}}) \leq r_\alpha^\varepsilon(XY) \ ,$$

which concludes the proof. □

In particular, the Rényi entropy of a random variable Z cannot increase when applying a function f to Z, that is,

$$H_\alpha(Z) \geq H_\alpha(f(Z)) \tag{16}$$

for $\alpha \in \mathbb{R}^\infty$.

6.2.5 Independent random variables and Markov chains

Lemma 6.7. *Let X and Y be two independent random variables and let $\varepsilon, \varepsilon' \in \mathbb{R}^+$. Then, for any parameter α,*

$$H_\alpha^\varepsilon(X|Y) = H_\alpha^\varepsilon(X) \ .$$

Proof. Let $\mathcal{E}$ be an event with $\mathrm{Prob}[\mathcal{E}] \geq 1-\varepsilon$ such that

$$r_\alpha(P_{X,\mathcal{E}}) \leq r_\alpha^\varepsilon(X) \ .$$

Then

$$r_\alpha^\varepsilon(X) \geq r_\alpha(P_{X,\mathcal{E}}) = \max_y r_\alpha(P_{X,\mathcal{E}|Y=y}) \geq r_\alpha^\varepsilon(X|Y) \ ,$$

where the equality holds since $(X,\mathcal{E})$ is independent of Y. On the other hand, let now $\mathcal{E}$ be an event with $\mathrm{Prob}[\mathcal{E}] \geq 1-\varepsilon$ such that

$$r_\alpha^\varepsilon(X|Y) \geq \max_y r_\alpha(P_{X,\mathcal{E}|Y=y}) \ .$$

Let $\bar{y}$ be the value such that $P_{\mathcal{E}|Y=\bar{y}}$ is maximal, and let $\mathcal{E}'$ be the event defined by $P_{\mathcal{E}'|X=x} := P_{\mathcal{E}|X=x,Y=\bar{y}}$. Then, $\mathrm{Prob}[\mathcal{E}'] \geq \mathrm{Prob}[\mathcal{E}] \geq 1-\varepsilon$, and, obviously

$$\max_y r_\alpha(P_{X,\mathcal{E}|Y=y}) \geq r_\alpha(P_{X,\mathcal{E}|Y=\bar{y}}) = r_\alpha(P_{X,\mathcal{E}'}) \ .$$

Hence,

$$r_\alpha^\varepsilon(X|Y) \geq r_\alpha(P_{X,\mathcal{E}'}) \geq r_\alpha^\varepsilon(X) \ ,$$

which concludes the proof. □

Corollary 6.8. *Let X, Y, and Z be random variables such that $X \leftrightarrow Z \leftrightarrow Y$ is a Markov chain and let $\varepsilon, \varepsilon' \in \mathbb{R}^+$. Then, for any parameter α,*

$$H_\alpha^\varepsilon(X|YZ) = H_\alpha^\varepsilon(X|Z) \ .$$

In particular,

$$H_\alpha^{\varepsilon+\varepsilon'}(XY|Z) \geq H_\alpha^\varepsilon(X|Z) + H_\alpha^{\varepsilon'}(Y|Z)$$

and

$$H_\alpha^{\varepsilon+\varepsilon'+\varepsilon''}(XYZ) \geq H_\alpha^\varepsilon(X|Z) + H_\alpha^{\varepsilon'}(YZ) \ .$$

Proof. Let ε_z, for $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$, such that $\mathbf{E}[\varepsilon_{y,z}] = \varepsilon$ and

$$H_\alpha^\varepsilon(X|YZ) \leq \min_{z\in\mathcal{Z}} H_\alpha^{\varepsilon_z}(X|Y, Z=z) \ .$$

Since, by assumption, $P_{XY|Z=z} = P_{X|Z=z} \times P_{Y|Z=z}$, for any $z \in \mathcal{Z}$, Lemma 6.7 implies

$$H_\alpha^\varepsilon(X|YZ) \leq \min_{z\in\mathcal{Z}} H_\alpha^{\varepsilon_z}(X|Y, Z=z) = \min_{z\in\mathcal{Z}} H_\alpha^{\varepsilon_z}(X|Z=z) \leq H_\alpha^\varepsilon(X|Z) \ .$$

To prove the converse inequality, let ε_z, for $z \in \mathcal{Z}$, such that $\mathbf{E}[\varepsilon_z] = \varepsilon$ and

$$H_\alpha^\varepsilon(X|Z) \leq \min_{z\in\mathcal{Z}} H_\alpha^{\varepsilon_z}(X|Z=z) \ .$$

Then, again from Lemma 6.7,

$$H_\alpha^\varepsilon(X|Z) \leq \min_{z\in\mathcal{Z}} H_\alpha^{\varepsilon_z}(X|Z=z) = \min_{z\in\mathcal{Z}} H_\alpha^{\varepsilon_z}(X|Y, Z=z) \leq H_\alpha^\varepsilon(X|YZ) \ .$$

The second and the third assertion then follow from Corollary 6.4. □

7 Applications

7.1 Randomness Extraction

The problem of extracting uniform randomness from a non-uniform source has first been studied in [7], and later been defined explicitly in [9]. Today, randomness extraction is a well-known and widely used concept of theoretical computer science, and, particularly, of cryptography. Basically, a (strong) extractor is a function f which takes as input a random variable Z and some additional uniformly distributed randomness R, such that, if Z satisfies a certain entropy condition, the output $f(Z, R)$ (together with the additional randomness R) is almost uniformly distributed.

Definition 7.1. A *strong* $(\tau, \kappa, \varepsilon)$-extractor on a set $\mathcal{Z}$ is a function with domain $\mathcal{Z} \times \mathcal{R}$ (for a set $\mathcal{R}$) and range $\mathcal{U}$ of size $|\mathcal{U}| = 2^\tau$, such that, for any random variable Z on $\mathcal{Z}$ satisfying $H_\infty(Z) \geq \kappa$ and R uniformly distributed over $\mathcal{R}$, $d(f(Z, R)|R) \leq \varepsilon$ holds.

The following result has originally been proven in [7] based on two-universal hash functions (where the randomness R is used to select a function from a class of two-universal hash functions). Later, similar statements have been shown in [1] and [6].[3]

Lemma 7.2 (Leftover hash lemma). *For $\kappa > \tau$, there exists a strong $(\tau, \kappa, 2^{-(\kappa-\tau)/2})$-extractor.*

The following measure is closely related to *smooth entropy* as defined in [4] and [3]. For a set $\mathcal{P}$ of probability distributions, it quantifies the amount of uniform randomness which can be extracted from a random variable Z with probability distribution P_Z, where it is only known that P_Z is contained in $\mathcal{P}$ (in particular, the extractor does not depend on the actual distribution P_Z).

Definition 7.3. Let $\mathcal{P}$ be a set of probability distributions with range $\mathcal{Z}$. The *ε-extractable randomness* of $\mathcal{P}$ is

$$H^{\varepsilon}_{\mathrm{ext}}(\mathcal{P}) := \max_{\mathcal{U}:\, \exists f:\, \Gamma^{\mathcal{P}}_{\varepsilon}(\mathcal{Z}\to\mathcal{U})} \log(|\mathcal{U}|)$$

where $\Gamma^{\mathcal{P}}_{\varepsilon}(\mathcal{Z} \to \mathcal{U})$ denotes the set of functions f from $\mathcal{Z} \times \mathcal{R}$ (for some set $\mathcal{R}$) to $\mathcal{U}$ such that, for any random variable Z with $P_Z \in \mathcal{P}$ and R uniformly distributed over $\mathcal{R}$, $d(f(Z,R)|R) \leq \varepsilon$ holds.

As mentioned in the introduction, smooth Rényi entropy quantifies the amount of extractable uniform randomness, up to some small additive constant.

Theorem 7.4. *For any set $\mathcal{P}$ of probability distributions with range $\mathcal{Z}$ and $\varepsilon, \varepsilon_1, \varepsilon_2 \in \mathbb{R}^+$ with $\varepsilon_1 + \varepsilon_2 = \varepsilon$,*

$$\min_{P\in\mathcal{P}}(H^{\varepsilon_1}_{\infty}(P)) + 2\log(\varepsilon_2) \leq H^{\varepsilon}_{\mathrm{ext}}(\mathcal{P}) \leq \min_{P\in\mathcal{P}}(H^{\varepsilon}_{\infty}(P)) \ .$$

Proof. We start proving the first inequality. Let $\delta > 0$, let $n = H^{\varepsilon_1}_{\infty}(P) - \delta$, $r = n - \log \varepsilon_1$, and let F be a (n, r, ε_2)-strong extractor, which exists according to Lemma 7.2.

Let now $P \in \mathcal{P}$, let W be a random variable with $P_W = P$, and let Z be a random variable with

$$\delta(Z, W) \leq \varepsilon_1$$

and

$$H_{\infty}(Z) \geq n \ .$$

By the definition of strong extractors, we have

$$d(F(Z)|F) \leq \varepsilon_2 \ . \tag{17}$$

Let W be a random variable with $P_W = P$ and let U be a random variable uniformly distributed over the range of F. Using (1) and the triangle inequality, we have, for any function f,

$$d(f(W)) = \delta(f(W), U) \leq d(f(W), f(Z)) + \delta(f(Z), U) \leq d(W, Z) + \delta(f(Z), U) \leq \varepsilon_1 + d(f(Z)) \ .$$

Applying this to (17), we obtain

$$d(F(Z)|F) \leq \varepsilon \ .$$

Since this holds for all $P \in \mathcal{P}$, we conclude

$$H^{\varepsilon}_{\mathrm{ext}}(\mathcal{P}) \geq r = \min_{P\in\mathcal{P}}(H^{\varepsilon/2}_{\infty}(P)) - \log \varepsilon/2 - \delta \ .$$

The first inequality of Theorem 7.4 then follows from the fact that this holds for any $\varepsilon > 0$.

[3]For a simple proof of Lemma 7.2, see, e.g., [8], p. 20.

Let us now show the second inequality. Let F be a random function such that for all $P_Z \in \mathcal{P}$, $d(F(Z)|F) \leq \varepsilon$ holds. It suffices to verify that $\log|\mathcal{R}_F| \leq \min_{P\in\mathcal{P}}(H_\infty^\varepsilon(P))$. Indeed, we conclude from (16) that, for any $P_Z \in \mathcal{P}$,

$$H_\infty^\varepsilon(Z) \geq H_\infty^\varepsilon(f(Z)) = \log|\mathcal{R}_F| \ ,$$

where the second equality follows from the assumption $d(F(Z)|F) \leq \varepsilon$ and ??. □

In particular, using Lemma 4.2, we can conclude that the Rényi entropy of order α, for $\alpha > 1$, is a lower bound for the number of uniform random bits that can be extracted, i.e.,

$$\min_{P\in\mathcal{P}}(H_\alpha(P)) + \frac{\log(\varepsilon/2)}{\alpha - 1} + 2\log(\varepsilon/2) \leq H_{\mathrm{ext}}^\varepsilon(\mathcal{P}) \ .$$

7.2 Compression and Error Correction

Another fundamental property of a probability distribution P is is the minimum length of an encoding $C = E(Z)$ of a random variable Z with $P_Z = P$ such that Z can be retrieved from C with high probability. A similar quantity can be defined for a set $\mathcal{P}$ of probability distributions. As a motivating example, consider the following setting known as *information reconciliation* [2]. An entity (Alice) holds a value Z which she wants to transmit to another entity (Bob), using τ bits of communication C. Clearly the minimum number τ of communication bits needed depends on the initial knowledge of Bob, which might be specified by some additional random variable W (not necessarily known to Alice). From Bob's point of view, the random variable Z is thus initially distributed according to $P_{Z|W=w}$, for some $w \in \mathcal{W}$. Consequently, in order to guarantee that Bob can reconstruct the value of Z with high probability, the error correcting information C sent by Alice must be useful for most of the distributions $P_{Z|W=w}$.

For the following, note that any probabilistic encoding function E corresponds to a deterministic functions e taking some additional randomness R, i.e., $E(Z) = e(Z, R)$.

Definition 7.5. A $(\tau, \kappa, \varepsilon)$-*encoding* on a set $\mathcal{Z}$ is a pair of functions (e, g) together with a random variable R with range $\mathcal{R}$ where e, the *encoding function*, is a mapping from $\mathcal{Z} \times \mathcal{R}$ to $\mathcal{C}$, for some set $\mathcal{C}$ of size $|\mathcal{C}| = 2^\tau$, and g, the *decoding function*, is a mapping from $\mathcal{C} \times \mathcal{R}$ to $\mathcal{Z}$ such that, for any random variable Z with range $\mathcal{Z}$ satisfying $H_0(Z) \leq \kappa$, $\mathrm{Prob}[g(e(Z, R), R) \neq Z] \leq \varepsilon$ holds.

The following result originally has been shown in the context of information reconciliation [2].

Lemma 7.6. *For $\tau > \kappa$, there exists a $(\tau, \kappa, 2^{-(\tau-\kappa)})$-encoding.*

For a set $\mathcal{P}$ of probability distributions, the measure defined below quantifies the minimum length of an encoding $C = e(Z, R)$ of Z such that Z can be reconstructed from C and R (with high probability), for any $P_Z \in \mathcal{P}$.

Definition 7.7. Let $\mathcal{P}$ be a set of probability distributions with range $\mathcal{Z}$. The ε-*encoding length* of $\mathcal{P}$ is

$$H_{\mathrm{enc}}^\varepsilon(\mathcal{P}) := \min_{\mathcal{C}:\, \exists e:\, \Lambda_\varepsilon^{\mathcal{P}}(\mathcal{Z}\to\mathcal{C})} \log(|\mathcal{C}|)$$

where $\Lambda_\varepsilon^{\mathcal{P}}(\mathcal{Z} \to \mathcal{C})$ denotes the set of pairs (e, R) where R is a random variable with range $\mathcal{R}$ and e a function from $\mathcal{Z} \times \mathcal{R}$ to $\mathcal{C}$ such that, for any random variable Z with $P_Z \in \mathcal{P}$, there exists a decoding function g from $\mathcal{C} \times \mathcal{R}$ to $\mathcal{Z}$ with $\mathrm{Prob}[g(e(Z, R), R) \neq Z] \leq \varepsilon$.

Similar the the amount of extractable randomness, smooth Rényi entropy can be used to characterize the minimum encoding length.

Theorem 7.8. *For any set $\mathcal{P}$ of probability distributions with range $\mathcal{Z}$ and $\varepsilon, \varepsilon_1, \varepsilon_2 \in \mathbb{R}^+$ with $\varepsilon_1 + \varepsilon_2 = \varepsilon$,*

$$\max_{P\in\mathcal{P}}(H_0^\varepsilon(P)) \leq H_{\mathrm{enc}}^\varepsilon(\mathcal{P}) \leq \max_{P\in\mathcal{P}}(H_0^{\varepsilon_1}(P)) - \log(\varepsilon_2) \ .$$

8 Conclusion

We have introduced a new class of entropy measures, called *smooth Rényi entropy* of order α. For $\alpha = 0$ and $\alpha = \infty$, the smooth Rényi entropy of a random variable Z corresponds to fundamental properties of Z, namely the minimum length of an encoding of Z and the amount of uniform randomness that can be extracted from Z, respectively. Moreover, if Z consists of n independent realizations of the same random experiment, the smooth Rényi entropy of Z approaches the Shannon entropy of Z, as n goes to infinity. Shannon entropy can thus be interpreted as the common limes of two different fundamental measures of entropy, for the (important) special case of many independent realizations.

References

[1] C. H. Bennett, G. Brassard, C. Crépeau, and U. Maurer. Generalized privacy amplification. IEEE Transactions on Information Theory, 41(6):1915-1923, 1995.

[2] G. Brassard and L. Salvail. Secret-key reconciliation by public discussion. In Springer Verlag, editor, *Adv. of Cryptology — EUROCRYPT'93*, Lecture Notes in Computer Science, pages 410–423, 1994.

[3] C. Cachin Smooth entropy and Rényi entropy. In *Adv. in Cryptology — EUROCRYPT '97*, volume 1233 of *Lecture Notes in Computer Science*, pages 193–208. Springer-Verlag, 1997.

[4] C. Cachin and U. Maurer. Smoothing probability distributions and smooth entropy. In *Proceedings of ISIT*, 1997.

[5] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, 1991.

[6] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computation*, 28(4):1364–1396, 1999.

[7] R. Impagliazzo, L. A. Levin, and M. Luby. Pseudo-random generation from one-way functions (extended abstract). In *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*, pages 12–24, 1989.

[8] M. Luby and A. Wigderson. Pairwise independence and derandomization. Technical Report CSD-95-880, Computer Science Institute, Berkeley, CA, 1995. `http://citeseer.nj.nec.com/luby95pairwise.html`.

[9] D. Zuckerman and N. Nisan. Randomness is linear in space. *Journal of Computer and System Sciences*, 52:43–52, 1996. A preliminary version appeared at STOC'93.

[10] A. Rényi. On measures of entropy and information. In *Proceedings of the 4th Berkeley Symp. on Math. Statistics and Prob.*, volume 1, pages 547–561. Univ. of Calif. Press, 1961.