Stefan Berens

Conditional Rényi entropy

Master thesis, defended on 28 August 2013

Thesis advisor:
Serge Fehr, CWI Amsterdam
Richard Gill, Universiteit Leiden

Specialisation:
Algebra, Geometry and Number Theory

Mathematisch Instituut, Universiteit Leiden

Acknowledgement

I would like to express my deepest gratitude to my advisor Serge Fehr for the extensive support of the project that resulted in this thesis. The guidance he offered me throughout this project was invaluable to me. Especially, I want to thank him for his patience, motivation, enthusiasm and immense knowledge that benefitted the project tremendously.

I would also like to extend my sincere gratitude to my co-advisor Richard Gill for the additional support.

It is safe to say that this thesis would have not been possible without the support and help of these two people.

Leiden, 14 August 2013 — Stefan Berens

Abstract

The introduction of the Rényi entropy allowed a generalization of the Shannon entropy and unified its notion with that of other entropies. However, so far there is no generally accepted conditional version of the Rényi entropy corresponding to the one of the Shannon entropy. Different definitions proposed so far in the literature lacked central and natural properties one way or another.

In this thesis we propose a new definition for the conditional case of the Rényi entropy. Our new definition satisfies all of the properties we deem natural. First and foremost, it is consistent with the existing, commonly accepted, definition of the conditional Shannon entropy as well as with the right notion of the conditional min entropy. Furthermore, and in contrast to previously suggested definitions, it satisfies the two natural properties that are monotonicity and (weak) chain rule and which we feel need to be satisfied by any 'good' entropy notion.

Another characteristic of our new definition is that it can be formulated in terms of the Rényi divergence. Additionally, it enables the use of (entropy) splitting. We conclude with an application where we use our new entropy notion as a tool to analyze a particular quantum cryptographic identification scheme.

Contents

1. Introduction: content of the thesis

In the article 'A Mathematical Theory of Communication', published in July 1948 in 'The Bell System Technical Journal', Claude Elwood Shannon first introduced his mathematical theory of information (cf. [11]). In particular, he presented the concept of entropy as a measure for information. Shannon's work represents the foundation of today's information theory.

On this foundation Alfréd Rényi then build one of his contributions, published in form of 'On Measures of Information and Entropy' in 1961 in 'Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960' (cf. [10]). At the center, he introduced a new notion of entropy of order α that included the one of Shannon as the special case $\alpha \to 1$.

In the original article, Shannon also defined the notion of conditional entropy, a measure of information in case of additional side information. However, a similar and generalized concept is not present in the paper of Rényi. On top of that, all definitions that have been proposed so far behave unnatural and undesirable in one or another way (for example analyzed in the recent survey article 'Conditional Rényi Entropies' by Antunes, Matos and Teixeira; cf. [1]).

The goal of the project that resulted in this thesis was to find the 'right' definition for the conditional Rényi entropy. That is to say, one that satisfies all the properties one would naturally expect from a 'good' entropy notion.

First and foremost, one would expect that an adequate definition for the conditional Rényi entropy generalizes the conditional entropy proposed by Shannon (analogous to the unconditional case). In addition, it should generalize another form of (conditional) entropy notion, the (conditional) min entropy. Furthermore, two other properties should be satisfied and will be of central interest, as they arise naturally and are not satisfied by previously suggested definitions. Namely, the entropy H_α of order $\alpha \in \mathbb{R}_{\geq 0}$ of a random variable X should only decrease when additional side information in form of a random variable Y is present, i.e. $H_\alpha(X) \geq H_\alpha(X|Y)$, which we will call 'monotonicity'. Moreover, it should do so by no more than $H_0(Y)$, the number of bits needed to represent Y, i.e. $H_\alpha(X|Y) \geq H_\alpha(X) - H_0(Y)$, which we will refer to as '(weak) chain rule'.

Additionally, we will show how the new conditional Rényi entropy actually follows rather naturally from the notion of Rényi divergence. Moreover, we will analyze its behavior with respect to the concept of (entropy) splitting. Using the latter, we will apply the new definition to the work of Damgård, Fehr, Salvail and Schaffner on a particular scheme of quantum cryptographic identification and thereby extend its security analysis (cf. [8]).

2. Preliminary

In this preliminary section we will introduce some basic concepts, as well as general conventions, specific notations and other useful tools. A reader that is already familiar with the topics of information theory and cryptography is likely accustomed to the content of this section. Nevertheless, we recommend looking into the conventions to avoid any confusions later on.

2.1. Logarithm.

To start with, we agree on:

Convention. In this thesis, $\log(x)$ will always refer to $\log_2(x)$, the 'logarithm of x to base 2'. In addition, we will adhere to the convention of writing $\ln(x)$ instead of $\log_e(x)$. Furthermore, we may write $\log x$ instead of $\log(x)$. The somewhat ambiguously notation of $\log x^y$ should be read as $\log(x^y)$, $\log(x+y)^z$ as $\log((x+y)^z)$, etc (i.e. the operation of exponentation has higher precedence than taking the logarithm).

2.2. Probability theory.

Probability theory will naturally be at the center of this thesis. A probability space is in general defined as follows:

Definition 1. *A probability space is defined to be a triple* $(\Omega, \mathcal{F}, P)$*, where* Ω *is an arbitrary non-empty set,* $\mathcal{F}$ *a* σ*-algebra over* Ω *and* $P\colon \mathcal{F} \to [0,1]$ *a probability measure.*

The subsets $\mathcal{A}$*,* $\mathcal{B}$ *of* Ω *are called events. If* $\mathcal{A}$ *is an event then* $P(\mathcal{A})$ *is referred to as 'probability of* $\mathcal{A}$*' and if* $\mathcal{A}$ *and* $\mathcal{B}$ *are events with* $P(\mathcal{B}) > 0$*, then the '(conditional) probability of* $\mathcal{A}$ *given* $\mathcal{B}$*' is defined to be* $P(\mathcal{A}|\mathcal{B}) := P(\mathcal{A} \cap \mathcal{B})/P(\mathcal{B})$*.*

However, in this thesis we will restrict ourselves to the finite case:

Definition 2. *A finite probability space is a probability space* $(\Omega, \mathcal{F}, P)$*, where* Ω *is finite and* $\mathcal{F} = \mathcal{P}(\Omega)$ *(the power set of* Ω*).*

In fact, in case of the finite probability space $(\Omega, \mathcal{P}(\Omega), P)$ one can simply write (Ω, P) with the fixed choice $\mathcal{F} = \mathcal{P}(\Omega)$ being understood.

Now, let us give the definition of a (probability) distribution:

Definition 3. *Let* $\mathcal{X}$ *be a finite non-empty set and let* $Q\colon \mathcal{X} \to \mathbb{R}_{\geq 0}$ *be a function. If* $\sum_{x\in\mathcal{X}} Q(x) = 1$ *and thus* $Q\colon \mathcal{X} \to [0,1]$*, then* Q *is called a (probability) distribution over* $\mathcal{X}$*. If* $\sum_{x\in\mathcal{X}} Q(x) \neq 1$*, then* Q *is called non-normalized distribution over* $\mathcal{X}$*.*

Next, follows a recap of the definition for a (finite) random variable:

Definition 4. *Let* (Ω, P) *be a fixed finite probability space and let* $\mathcal{X}$ *be a finite non-empty set. A function* $X\colon \Omega \to \mathcal{X}$ *is called a* $\mathcal{X}$*-valued (finite) random variable. The set* $\mathcal{X}$ *is called the set of values (of* X*) and its elements* $x \in \mathcal{X}$ *the values (of* X*). The associated probability distribution* $P_X\colon \mathcal{X} \to [0,1]$ *is given via* $P_X(x) := P(X^{-1}(x))$ *for* $x \in \mathcal{X}$ *with* $X^{-1}(x) \in \mathcal{P}(\Omega)$ *being the inverse image (of* x *under* X*).*

One specific type of probability distribution used here will be:

Definition 5. *Let $\mathcal{X}$ be a finite non-empty set. Define the uniform distribution over $\mathcal{X}$ as the probability distribution $U_{\mathcal{X}}\colon \mathcal{X} \to [0,1]$ given by $x \mapsto \frac{1}{|\mathcal{X}|}$.*

Similarly, the corresponding random variable:

Definition 6. *Let (Ω, P) be a fixed finite probability space and let $\mathcal{X}$ be a finite non-empty set. The $\mathcal{X}$-valued random variable X is called uniformly distributed over $\mathcal{X}$, if its associated probability distribution is identical to the uniform distribution over $\mathcal{X}$, i.e. it is given by $P_X(x) = \frac{1}{|\mathcal{X}|}$ for all $x \in \mathcal{X}$.*

In the next step, we will start to recall the case of more than one random variable (and their associated probability distribution):

Notation. Let (Ω, P) be a fixed finite probability space and let X, Y be $\mathcal{X}$- resp. $\mathcal{Y}$-valued random variables. The associated probability distribution of the $\mathcal{X} \times \mathcal{Y}$-valued random variable $(X,Y)\colon \Omega \to \mathcal{X} \times \mathcal{Y}$ given via $\omega \mapsto (X(\omega), Y(\omega))$ is denoted by $P_{X,Y}$ and referred to as the 'joint probability distribution of X and Y'.

The function $P_{X,Y}(\cdot, y)\colon \mathcal{X} \to [0,1]$, obtained by fixing the second argument of $P_{X,Y}$ to $y \in \mathcal{Y}$, is denoted by $P_{X,Y=y}$. Analogously, this is done when interchanging the roles of X and Y.

Now, we can give the conditional version of a probability distribution:

Definition 7. *Let (Ω, P) be a fixed finite probability space and let X, Y be $\mathcal{X}$- resp. $\mathcal{Y}$-valued random variables. If $y \in \mathcal{Y}$ such that $P_Y(y) > 0$, then define the conditional probability distribution of X given that Y is equal to y as $P_{X|Y=y}\colon \mathcal{X} \to [0,1]$ via $P_{X|Y=y}(x) := P_{X,Y}(x,y)/P_Y(y)$ for $x \in \mathcal{X}$, $y \in \mathcal{Y}$. Also, write $P_{X|Y}(x|y)$ for $P_{X|Y=y}(x)$.*

Furthermore, we use the statistical distance, defined as follows, as distance measure for probability distributions:

Definition 8. *Let $\mathcal{X}$ be a finite non-empty set and let Q_1 and Q_2 be probability distributions over $\mathcal{X}$. Then, the statistical distance between Q_1 and Q_2 is defined as*

$$\Delta[Q_1, Q_2] := \frac{1}{2} \sum_{x \in \mathcal{X}} |Q_1(x) - Q_2(x)|.$$

Note that, as Δ is a function of probability distributions, one can apply it to the associated probability distributions of random variables. Namely, if X_1 and X_2 are $\mathcal{X}$-valued random variables on a fixed finite probability space (Ω, P), then we apply Δ to P_{X_1} and P_{X_2}.

In the following, we will state some conventions that we use. First, we need the following definition:

Definition 9. *Let $\mathcal{X}$ be an arbitrary set and furthermore let $f\colon \mathcal{X} \to \mathbb{R}$ be an arbitray function. Define the support of f as*

$$\operatorname{supp}(f) := \{x \in \mathcal{X} \mid f(x) \neq 0\}.$$

Convention. For any finite probability space (Ω, P), we always assume that $P(\{\omega\}) > 0$ for all $\omega \in \Omega$.

Note, that this is without loss of generality as we can always replace Ω by $\operatorname{supp}(P)$. Additionally, from now on and throughout the thesis we leave the specific finite probability space implicit. Whenever we refer to a random variable X, we understand an arbitrary but fixed probability space to be given and as such the distribution P_X of X is given as well. Furthermore, in order to avoid expressions like $0/0$, we will use the following simplification:

Convention. For any $\mathcal{X}$-valued (finite) random variable X, we always assume that $P_X(x) > 0$ for all $x \in \mathcal{X}$.

Again, this is without loss of generality as we can (similar to before) always replace $\mathcal{X}$ by $\operatorname{supp}(P_X)$. In addition, let us state the following:

Convention. If not specified otherwise, the random variable X has the set of values $\mathcal{X}$, the random variable Y has the set of values $\mathcal{Y}$, etc.

2.3. **Jensen's inequality.**

In this thesis, we will extensively use the inequality proven by the Danish mathematician Johan Jensen in 1906, cf. [5], which we state as follows:

Theorem 2.1. *Let $\varphi\colon \mathbb{R} \to \mathbb{R}$ be a convex function and $n \in \mathbb{N}$. Then, for any $p_1, \ldots, p_n \in \mathbb{R}_{\geq 0}$ with $\sum_{i=1}^n p_i = 1$ and $x_1, \ldots, x_n \in \mathbb{R}$ it holds that*

$$\varphi\Big(\sum_{i=1}^n p_i \cdot x_i\Big) \leq \sum_{i=1}^n p_i \cdot \varphi(x_i);$$

An immediate consequence is, that if φ is instead concave, then it holds that

$$\varphi\Big(\sum_{i=1}^n p_i \cdot x_i\Big) \geq \sum_{i=1}^n p_i \cdot \varphi(x_i).$$

Also, if φ is in fact strictly convex (resp. concave) and $p_1, \ldots, p_n > 0$, then equality holds if and only if $x_1 = \ldots = x_n$.

Proof. Claim follows via straightforward induction on n. □

2.4. **The p-(quasi)norm.**

Finally, we quickly recall the concept of the p-(quasi)norm of a function (with finite domain):

Definition 10. *Let $f\colon \mathcal{X} \to \mathbb{R}$ be a function on a finite set $\mathcal{X}$. Define the p-norm, for $1 \leq p < \infty$, or p-quasinorm, for $0 < p < 1$, of f as:*

$$||f||_p := \Big(\sum_{x \in \mathcal{X}} |f(x)|^p \Big)^{\frac{1}{p}}$$

The ∞-norm, also called maximum norm, of f is defined as

$$||f||_\infty := \max_{x \in \mathcal{X}} \{|f(x)|\}.$$

It holds that $||f||_\infty = \lim_{p\to\infty} ||f||_p$ and thus one might wonder about the other limit, $\lim_{p\to 0} ||f||_p$. In some texts in the literature (e.g. [6]) the zero 'norm' of f is defined as

$$||f||_0 := |\operatorname{supp}(f)|$$

but we stress, that in general it does not satisfy $||f||_0 = \lim_{p\to 0} ||f||_p$. Furthermore, we do not make use of this notion of zero 'norm'.

Note that the term p-norm, for $1 \leq p \leq \infty$, is justified by the fact that it is indeed a norm in the mathematical sense. The analog applies to the term p-quasinorm, for $0 < p < 1$, which is justified by the fact that it is indeed a quasinorm in the mathematical sense. Meaning that the quotation marks in case of the zero 'norm' are meant to indicate the fact that it is not a norm (in the mathematical sense).

Considering two values for p, one should recall the following relation:

Remark. Let $f\colon \mathcal{X} \to \mathbb{R}$ be a function on a non-empty finite set $\mathcal{X}$. Then, the p-(quasi-)norm of f is monotonically decreasing in p, i.e. for $\infty \geq p_1 \geq p_2 > 0$ it holds that

$$||f||_{p_1} \leq ||f||_{p_2}.$$

3. Entropy

In this section we will give a brief overview over the well-known concept of Shannon entropy as well as some other entropy notions, including the Rényi entropy of order α. Additionally, previous approaches regarding the conditional Rényi entropy are discussed. Note that most statements given here are given without explicit proof; they can be found in the standard literature.

3.1. **Shannon entropy.** The following will be a recap of the basics of the concept of Shannon entropy. First, let us state its core definition:

Definition 11. *Let X be a $\mathcal{X}$-valued random variable with associated probability distribution P_X. Then, the Shannon entropy H of X is defined as*

$$H(X) := -\sum_{x\in\mathcal{X}} P_X(x)\log P_X(x).$$

We point out that the Shannon entropy H is actually a function of the associated probability distribution P_X of X. However, out of convenience, the common (abusive) notation $H(X)$ is used instead of $H(P_X)$. In addition, note that by convention one has $P_X(x) > 0$ for every $x \in \mathcal{X}$ and thus the expression $\log P_X(x)$ is always well defined. Alternatively, one could achieve this also by restricting the sum to terms with positive probability or equivalently define $0\log 0$ to be 0, which is justified by taking limits.

Now, let X and Y be two random variables. The resulting expression of applying H to the conditional probability distribution $P_{X|Y=y}$ for $y \in \mathcal{Y}$ is denoted by $H(X|Y = y)$ and allows the following:

Definition 12. *Let X, Y be two random variables respectively $\mathcal{X}$-, $\mathcal{Y}$-valued and with joint probability distribution $P_{X,Y}$. The (conditional) Shannon entropy H of X given Y is then defined as*

$$\begin{aligned} H(X|Y) :=& \sum_{y\in\mathcal{Y}} P_Y(y) H(X|Y=y) \\ =& -\sum_{y\in\mathcal{Y}} P_Y(y) \sum_{x\in\mathcal{X}} P_{X|Y}(x|y)\log P_{X|Y}(x|y). \end{aligned}$$

Note that, again by convention, $P_Y(y) > 0$ for every $y \in \mathcal{Y}$. Notice in addition the following lower and upper bound of the Shannon entropy, which - as we point out - are in fact optimal:

Proposition 3.1. *Let X be a random variable. The Shannon entropy (of X) satisfies the inequalities:*

$$0 \le H(X) \le \log|\mathcal{X}|$$

In addition, equality holds on the lower end of the chain of inequalities if and only if X is deterministic, i.e. $\mathcal{X} = \{x\}$ and thus $P_X(x) = 1$.

And in contrast, on the upper end equality holds if and only if X is uniformly distributed over $\mathcal{X}$, i.e. $P_X(x) = \frac{1}{|\mathcal{X}|}$ for all $x \in \mathcal{X}$.

Proof. The claim for the left side follows directly from the definition; The one for the right side follows by also using Jensen's inequality. □

Next, recall how the two natural properties that were first introduced in Section 1 correspond to an intuitive expectation and are in fact satisfied by the Shannon entropy.

Namely, the first one, monotonicity, corresponds to the intuition that uncertainty can only drop with additional side information being present. On the other hand, the second one, chain rule, captures the intuition that the uncertainty can not drop by more than the amount of additional information that is present. In addition, it holds that:

Proposition 3.2. *Let X, Y, Z be random variables. Then, for the (conditional) Shannon entropy the following holds:*

i) Monotonicity

$$H(X|Z) \geq H(X|Y, Z)$$

ii) Chain rule

$$H(X|Y, Z) \geq H(X, Y|Z) - \log|\mathcal{Y}| \geq H(X|Z) - \log|\mathcal{Y}|$$

Notice that the 'chain rule' stated above is actually a weaker version of what is usually referred to as the 'chain rule for Shannon entropy' (cf. [2]):

Remark. Let X, Y, Z be all random variables. The Shannon entropy satisfies the chain rule in its stronger version, namely

$$H(X, Y|Z) = H(X|Y, Z) + H(Y|Z)$$

which implies the regular one by using Proposition 3.1 and 3.2 i).

3.2. **Beyond Shannon entropy.** In addition to the Shannon entropy, there are some other important entropy notions, which we briefly recall here.

3.2.1. *Min entropy.* First of all, we will begin with the definition of the so-called min entropy and extend it then to the conditional case. Note that we abuse the notation with respect to the dependence in the same way we did for the Shannon entropy: The min entropy is actually a function of the probability distribution but we will instead write it as a function of the random variable.

Definition 13. *Let X be a random variable. Define the min entropy H_∞ of X as follows*

$$H_\infty(X) := -\log \operatorname{Guess}(X)$$

where the guessing probability of X is given by

$$\mathrm{Guess}(X) := \max_{x\in\mathcal{X}}\{P_X(x)\}.$$

The name 'guessing probability' steems from the fact that when given a random variable the highest probability of guessing the value is the maximum of the probabilities. Similar to the Shannon entropy one can consider again a conditional version (cf. [3]):

Definition 14. *Let X, Y be two random variables. The (conditional) min entropy H_∞ of X given Y is defined as*

$$H_\infty(X|Y) := -\log \mathrm{Guess}(X|Y)$$

where the (conditional) guessing probability of X given Y is defined as

$$\mathrm{Guess}(X|Y) := \sum_{y\in\mathcal{Y}} P_Y(y)\,\mathrm{Guess}(X|Y=y).$$

We point out that other definitions for the conditional min entropy have been proposed in the past. However, none of them actually satisfied both monotonicity and chain rule (which, as was shown, arise naturally). On the other hand, this one here does:

Proposition 3.3. *Let X, Y, Z be random variables. Then, for the (conditional) min entropy the following holds:*

i) Monotonicity

$$H_\infty(X|Z) \geq H_\infty(X|Y,Z)$$

ii) Chain rule

$$H_\infty(X|Y,Z) \geq H_\infty(X,Y|Z) - \log|\mathcal{Y}| \geq H_\infty(X|Z) - \log|\mathcal{Y}|$$

Since this is a rather new notion of conditional min-entropy and as such not covered in standard text books, we prove Proposition 3.3 for completeness:

Proof. For simplicity we only prove the case of an empty Z at this point. The proof for monotonicity then goes as follows:

$$\begin{aligned}
\mathrm{Guess}(X) &= \max_{x\in\mathcal{X}}\{P_X(x)\} \\
&= \max_{x\in\mathcal{X}}\Big\{\sum_{y\in\mathcal{Y}} P_Y(y)P_{X|Y}(x|y)\Big\} \\
&\leq \sum_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}}\{P_Y(y)P_{X|Y}(x|y)\} \\
&= \sum_{y\in\mathcal{Y}} P_Y(y)\max_{x\in\mathcal{X}}\{P_{X|Y}(x|y)\} \\
&= \mathrm{Guess}(X|Y)
\end{aligned}$$

The (in-)equalities should be self-explanatory. Almost as straightforward is the derivation of the chain rule:

$$\begin{aligned}
\mathrm{Guess}(X|Y) &= \sum_{y\in\mathcal{Y}} P_Y(y) \max_{x\in\mathcal{X}}\{P_{X|Y}(x|y)\} \\
&\leq \max_{y\in\mathcal{Y}} \left\{ P_Y(y) \max_{x\in\mathcal{X}}\{P_{X|Y}(x|y)\} \right\} |\mathcal{Y}| \\
&= \max_{x\in\mathcal{X},y\in\mathcal{Y}} \{P_{X,Y}(x,y)\}|\mathcal{Y}| \\
&= \mathrm{Guess}(X,Y)|\mathcal{Y}| \\
&= \max_{x\in\mathcal{X},y\in\mathcal{Y}} \{P_{X,Y}(x,y)\}|\mathcal{Y}| \\
&\leq \max_{x\in\mathcal{X},y\in\mathcal{Y}} \{P_X(x)\}|\mathcal{Y}| \\
&= \max_{x\in\mathcal{X}} \{P_X(x)\}|\mathcal{Y}| \\
&= \mathrm{Guess}(X)|\mathcal{Y}|
\end{aligned}$$

□

3.2.2. *Max entropy.* In the next step, we will recall and analyze the so-called max entropy. We point out that, as far as we know, there is no generally accepted conditional version for the max entropy.

Definition 15. *Let X be a random variable. Define the* <u>*max entropy*</u> *H_0 of X as follows*

$$H_0(X) := \log|\operatorname{supp}(P_X)|.$$

Remark. By our convention that $\mathcal{X} = \operatorname{supp}(P_X)$, we may also write

$$H_0(X) = \log|\mathcal{X}|.$$

and as such the maximal value of the Shannon entropy, as discussed in Proposition 3.1, is equal to the max entropy.

As there is no conditional version, it does not make sense to talk about monotonicity and (weak) chain rule, yet. Instead, we will show the so-called property of subadditivity:

Proposition 3.4. *Let X, Y be random variables. The max entropy satisfies subadditivity, i.e.*

$$H_0(X,Y) \leq H_0(X) + H_0(Y).$$

Proof. The obvious fact for the support sizes

$$|\operatorname{supp}(P_{X,Y})| \leq |\operatorname{supp}(P_X)| \cdot |\operatorname{supp}(P_Y)|$$

immediately yields the claim for the entropies. □

3.2.3. *Collision entropy.* Finally, we define the collision entropy. Note, that as before there is no generally accepted conditional version.

Definition 16. *Let X be a random variable. Then, define the collision entropy H_2 of X as*

$$H_2(X) := -\log(\mathrm{Col}(X)).$$

where the collision probability of X is given by

$$\mathrm{Col}(X) := \sum_{x \in \mathcal{X}} P_X(x)^2.$$

The expression 'collision probability of a random variable X' is due to the following interpretation. Let X' be another random variable with identical associated probability distribution as X but independent of it. In this case, the probability of X and X' colliding, i.e. of yielding the same value, is equal to the expression

$$\sum_{x \in \mathcal{X}} P_{X,X'}(x,x) = \sum_{x \in \mathcal{X}} P_X(x) P_{X'}(x) = \sum_{x \in \mathcal{X}} P_X(x)^2.$$

Notice, that we used the independence (of X and X') first and then the fact that the associated probability distributions are the same.

Now, the collision entropy is used for example in the context of the privacy amplification theorem, which we will also extend later on. Prior to this, we recall the notion of a universal hash function:

Definition 17. *Let $\mathcal{S}, \mathcal{X}, \mathcal{Z}$ be non-empty finite sets, let S be a random variable uniformly distributed over $\mathcal{S}$ and let $g \colon \mathcal{X} \times \mathcal{S} \to \mathcal{Z}$ be a fixed function. Then, g is called a universal hash function, if*

$$P(g(x,S) = g(x',S)) \leq \frac{1}{|\mathcal{Z}|}$$

for any choices $x \neq x'$ (both elements of $\mathcal{X}$).

Equipped with this definition, it is possible to state the previously mentioned privacy amplification theorem:

Theorem 3.5. *Let X, S be independent random variables where S is uniformly distributed over $\mathcal{S}$. Let further $g \colon \mathcal{X} \times \mathcal{S} \to \{0,1\}^r$ be a universal hash function, where $1 \leq r < \infty$, and define $K := g(X,S)$. Then,*

$$\Delta[P_{K,S}; U_{\{0,1\}^r} \cdot P_S] \leq \frac{1}{2} \cdot 2^{-\frac{1}{2}(H_2(X)-r)}.$$

Proof. The result was first published in [4] using the min entropy, but is easily extended to the conditional entropy. □

The privacy amplification theorem allows the following use case. First, assume you generated an n-bit string, the random variable X. Now, using a generated seed, the random variable S, and furthermore a universal hash function, the function g, you then want to compute a secure key K of length r by applying g to X and S. Provided the public knowledge of both the function g and the random variable S, the question is whether the key is in fact secure. The theorem now states that, as long as the n-bit string has significantly more than r bit of entropy, the generated key appears to be chosen uniformly.

In the next step, assume additional knowledge about the generated n-bit string, in form of the random variable Y. The question is whether the key is still secure. However, we will only be able to answer it later on in this thesis using the (then defined) conditional collision entropy.

3.2.4. *Relation.* All of the previously introduced entropies are related as follows:

Proposition 3.6. *Let X be a random variable. The different entropies can be ordered (and bounded above and below) in the following way:*

$$0 \leq H_\infty(X) \leq H_2(X) \leq H(X) \leq H_0(X) = \log|\mathcal{X}|$$

Proof. The proof is a straight-forward computation using, among other things, Jensen's inequality. □

3.3. **Rényi entropy.** Up to this point we have seen four entropies (including the original Shannon entropy). In the next step, we are going to recall the Rényi entropy as introduced by Rényi in his paper [10]. It unifies all entropies that we have introduced before:

Definition 18. *Let X be a random variable. Then, the Rényi entropy H_α of X of order $\alpha \in [0,1) \cup (1,\infty)$ is defined as*

$$H_\alpha(X) := -\log(\mathrm{Ren}_\alpha(X))$$

where the Rényi probability of X of order α is defined as

$$\mathrm{Ren}_\alpha(X) : = \Big(\sum_{x\in\mathcal{X}} P_X(x)^\alpha\Big)^{\frac{1}{\alpha-1}}.$$

By writing out the expression, we get what is usually used to introduce and define the Rényi entropy:

Remark. Let X be a random variable and $\alpha \in [0,1) \cup (1,\infty)$. Then,

$$H_\alpha(X) = \frac{1}{1-\alpha}\log\Big(\sum_{x\in\mathcal{X}} P_X(x)^\alpha\Big).$$

Furthermore, we can also use the concept of the p-(quasi)-norm:

Remark. For $\alpha \in (0,1) \cup (1,\infty)$ and a given random variable X the Rényi probability can also be written as

$$\mathrm{Ren}_\alpha(X) = ||P_X||_\alpha^{\frac{\alpha}{\alpha-1}}.$$

In the definition before we avoided the values of 1 and ∞ for α because of obvious reasons. Nonetheless, as noted in the next proposition below, those can be incorporated by taking limits. Additionally, the same proposition also shows that there is no clash in notation between the Rényi entropy and the previously introduced entropies as Rényi in fact generalized all previous entropies:

Proposition 3.7. *Let X be a random variable. Then, the two limits $\lim_{\alpha\to 1} H_\alpha(X)$ and $\lim_{\alpha\to\infty} H_\alpha(X)$ exist, and*

$$\begin{aligned} H_1(X) := \lim_{\alpha\to 1} H_\alpha(X) &= H(X) \textit{ and} \\ \lim_{\alpha\to\infty} H_\alpha(X) &= H_\infty(X). \end{aligned}$$

Furthermore, the Rényi entropy of order $\alpha = 2$ and $\alpha = 0$ coincides respectively with the introduced collision entropy (Definition 16) and max entropy (Definition 15).

Proof. A proof for the part about the generalization of Shannon can for example be found in the original work, [10]. Consult standard literature for the proofs of the other limits and identities. □

Furthermore, the exact relation of the different entropies given in Proposition 3.6 generalizes as follows:

Proposition 3.8. *Let X be a random variable. Now, if $\alpha, \beta \in [0, \infty]$ with $\alpha \geq \beta$, then $0 \leq H_\alpha(X) \leq H_\beta(X) \leq \log|\mathcal{X}|$.*

Proof. The first and last inequality are clear from the definitions. In addition, it is enough to consider $\alpha, \beta \in (0,1)\cup(1,\infty)$ as the other cases follow easily by using corresponding limits. Now, let $\alpha, \beta \in (1, \infty)$, then the claim is equivalent to

$$\Big(\sum_{x\in\mathcal{X}} P_X(x)^\alpha\Big)^{\frac{1}{\alpha-1}} \overset{!}{\geq} \Big(\sum_{x\in\mathcal{X}} P_X(x)^\beta\Big)^{\frac{1}{\beta-1}}.$$

Using the fact that $\alpha \geq \beta$, implying $\frac{(\beta-1)}{(\alpha-1)} \leq 1$, yields in combination with Jensen's inequality

$$\begin{aligned} \Big(\sum_{x\in\mathcal{X}} P_X(x)^\alpha\Big)^{\frac{1}{\alpha-1}} &= \Big(\sum_{x\in\mathcal{X}} P_X(x)P_X(x)^{\alpha-1}\Big)^{\frac{\beta-1}{(\alpha-1)(\beta-1)}} \\ &\geq \Big(\sum_{x\in\mathcal{X}} P_X(x)P_X(x)^{\frac{(\alpha-1)(\beta-1)}{\alpha-1}}\Big)^{\frac{1}{\beta-1}} \\ &= \Big(\sum_{x\in\mathcal{X}} P_X(x)^\beta\Big)^{\frac{1}{\beta-1}} \end{aligned}$$

This finishes the case $\alpha, \beta \in (1, \infty)$. On the other hand, the case of $\alpha, \beta \in (0, 1)$ follows by similar arguments. Finally, the case where $\alpha \in (1, \infty)$ and $\beta \in (0, 1)$ follows by transitivity. □

3.4. **Conditional Rényi entropy: previous approaches.** Similar to the case of the collision entropy and max entropy, there is as of now no commonly accepted definition for the conditional Rényi entropy. Thus, we give here a brief overview over the different suggestions we could find in the literature.

Let X, Y be random variables. A natural suggestion, which is similar to the approach for the conditional Shannon entropy, is

$$H_\alpha^1(X|Y) := \sum_{y\in\mathcal{Y}} P_Y(y) H_\alpha(X|Y=y)$$

which is discussed in the recent survey article [1] together with the following

$$\begin{aligned} H_\alpha^2(X|Y) :=& H_\alpha(X,Y) - H_\alpha(Y) \\ H_\alpha^3(X|Y) :=& \frac{1}{1-\alpha} \log \Big(\max_{y\in\mathcal{Y}} \Big\{ \sum_{x\in\mathcal{X}} P_{X|Y}(x|y)^\alpha \Big\} \Big) \end{aligned}$$

where H_α^2 is inspired by the (strong) chain rule for the Shannon entropy, i.e. $H(X,Y) = H(X|Y) + H(Y)$.

In the article [12] one can find another proposal, namely:

$$H_\alpha^4(X|Y) := \frac{1}{1-\alpha} \log \Big(\sum_{y\in\mathcal{Y}} P_Y(y) \sum_{x\in\mathcal{X}} P_{X|Y}(x|y)^\alpha \Big)$$

Another approach, without appearance in the literature albeit similar in its construction to the the conditional max entropy, is

$$H_\alpha^5(X|Y) := -\log(\mathrm{Ren}_\alpha(X|Y))$$

with $\mathrm{Ren}_\alpha(X|Y) := \sum_{y\in\mathcal{Y}} P_Y(y) \,\mathrm{Ren}_\alpha(X|Y=y)$.

For easier comparison, we write out all the previous suggestions:

$$\begin{aligned} H_\alpha^1(X|Y) &= \frac{1}{1-\alpha} \sum_{y\in\mathcal{Y}} P_Y(y) \log \left(\frac{\sum_{x\in\mathcal{X}} P_{Y,X}(y,x)^\alpha}{P_Y(y)^\alpha} \right) \\ H_\alpha^2(X|Y) &= \frac{1}{1-\alpha} \log \left(\frac{\sum_{y\in\mathcal{Y},x\in\mathcal{X}} P_{Y,X}(y,x)^\alpha}{\sum_{y\in\mathcal{Y}} P_Y(y)^\alpha} \right) \\ H_\alpha^3(X|Y) &= \frac{1}{1-\alpha} \log \left(\max_{y\in\mathcal{Y}} \left\{ \frac{\sum_{x\in\mathcal{X}} P_{X,Y}(x,y)^\alpha}{P_Y(y)^\alpha} \right\} \right) \\ H_\alpha^4(X|Y) &= \frac{1}{1-\alpha} \log \left(\sum_{y\in\mathcal{Y}} P_Y(y) \frac{\sum_{x\in\mathcal{X}} P_{X,Y}(x,y)^\alpha}{P_Y(y)^\alpha} \right) \\ H_\alpha^5(X|Y) &= \frac{1}{1-\alpha} \log \left(\sum_{y\in\mathcal{Y}} P_Y(y) \left(\frac{\sum_{x\in\mathcal{X}} P_{X,Y}(x,y)^\alpha}{P_Y(y)^\alpha} \right)^{\frac{1}{\alpha-1}} \right)^{\alpha-1} \end{aligned}$$

As we have already stressed before, none of the suggested definitions became commonly accepted. We believe this is to a large extent due to the following:

Remark. All above definitions of the conditional Rényi entropy do not satisfy both monotonicity and (weak) chain rule while simultaneously being a generalization of the (conditional) Shannon as well as of the (conditional) min entropy.

The proposals discussed in [1], i.e. H^1_α, H^2_α and H^3_α, lack the property of monotonicity, which is for example shown in the recent survey article. Moreover, as one can easily see, neither one of the other proposals, i.e. H^4_α and H^5_α, nor H^3_α satisfy the (weak) chain rule.

Concerning the second part of the remark, we point out that none of the definitions given above except H^5_α is consistent with the notion of the (conditional) min entropy $H_\infty(X|Y)$. Further, neither H^3_α nor H^5_α are consistent with the (conditional) Shannon entropy $H(X|Y)$.

4. Conditional Rényi entropy

In this section, we now propose a new definition of the conditional Rényi entropy. In contrast to previously suggested definitions, our (new) definition is consistent with the conditional Shannon entropy and conditional min entropy, and satisfies monotonicity and chain rule. Furthermore, in this section we cover the generalized relation between the conditional Rényi entropy of different orders. Another point of our analysis is the relation to the Rényi divergence. Finally, we touch on a concept called (entropy) splitting and how it holds for our definition.

4.1. Definition.

First, similar to the definition of the Rényi entropy, we will not define the conditional Rényi entropy for all values of α but rather use limits to take care of boundaries and gaps. Furthermore, recall that in the previous section we reformulated the Rényi entropy as minus the logarithm of the Rényi probability and we will use the very same approach here:

Definition 19. *Let X, Y be two random variables. Define the conditional Rényi entropy H_α of X given Y of order $\alpha \in (0,1) \cup (1,\infty)$ as*

$$H_\alpha(X|Y) := -\log(\mathrm{Ren}_\alpha(X|Y))$$

where the conditional Rényi probability of X given Y of order α is given by

$$\mathrm{Ren}_\alpha(X|Y) := \Big(\sum_{y\in\mathcal{Y}} P_Y(y)(\mathrm{Ren}_\alpha(X|Y=y))^{\frac{\alpha-1}{\alpha}}\Big)^{\frac{\alpha}{\alpha-1}}.$$

By writing out the conditional Rényi entropy, we obtain:

Remark. Let X, Y be two random variables, $\alpha \in (0,1) \cup (1,\infty)$. Then,

$$H_\alpha(X|Y) = -\log\Big(\sum_{y\in\mathcal{Y}} P_Y(y)\Big(\sum_{x\in\mathcal{X}} P_{X|Y}(x|y)^\alpha\Big)^{\frac{1}{\alpha}}\Big)^{\frac{\alpha}{\alpha-1}}.$$

Similar to the unconditional version, one can also reformulate the conditional Rényi probability using the α-(quasi-)norm:

Remark. Let X, Y be two random variables, $\alpha \in (0,1) \cup (1,\infty)$. Then,

$$\mathrm{Ren}_\alpha(X|Y) = \Big(\sum_{y\in\mathcal{Y}} P_Y(y)||P_{X|Y=y}||_\alpha\Big)^{\frac{\alpha}{\alpha-1}}.$$

4.2. Special cases.

In this subsection we will consider some important cases for the parameter α similar to the analysis of the Rényi entropy in Proposition 3.7. In particular, we will prove the claimed equality with the conditional Shannon as well as min entropy. Furthermore, we will give the consequential definition of the conditional max entropy. At last, we will state the then defined conditional collision entropy.

First of all, we will show consistency of our new definition in case of the limit of α going to 1 with the conditional Shannon entropy:

Proposition 4.1. *Let X, Y be random variables. Then,*

$$H_1(X|Y) := \lim_{\alpha\to 1} H_\alpha(X|Y) = H(X|Y).$$

Proof. First, for every $\alpha \in (0,1)\cup(1,\infty)$ one can re-formulate one side

$$\begin{aligned} H_\alpha(X|Y) &= -\log\Big(\sum_{y\in\mathcal{Y}} P_Y(y)||P_{X|Y=y}||_\alpha\Big)^{\frac{\alpha}{\alpha-1}} \\ &= -\frac{1}{1-\frac{1}{\alpha}}\log\Big(\sum_{y\in\mathcal{Y}} P_Y(y)||P_{X|Y=y}||_\alpha\Big) \\ &= -\frac{f(\alpha)}{g(\alpha)} \end{aligned}$$

where f and g are continuous functions defined for $\alpha \in (0,\infty)$ via:

$$\begin{aligned} f(\alpha) &:= \log\Big(\sum_{y\in\mathcal{Y}} P_Y(y)||P_{X|Y=y}||_\alpha\Big) \\ g(\alpha) &:= 1 - \frac{1}{\alpha} \end{aligned}$$

Note, that by continuity of the involved functions

$$\lim_{\alpha\to 1} f(\alpha) = f(1) = 0 \qquad\qquad \lim_{\alpha\to 1} g(\alpha) = g(1) = 0$$

and in conclusion $\lim_{\alpha\to 1}\frac{f(\alpha)}{g(\alpha)} = \frac{0}{0}$. Using L'Hospital's rule, this yields

$$\lim_{\alpha\to 1} H_\alpha(X|Y) = -\frac{\lim_{\alpha\to 1} f'(\alpha)}{\lim_{\alpha\to 1} g'(\alpha)}$$

under the assumption that the right hand side exists.

Furthermore, notice that $g'(\alpha) = \frac{1}{\alpha^2}$ and $f'(\alpha) = \frac{h'(\alpha)}{h(\alpha)\ln(2)}$ for the continuous function h with continuous derivative h', which are both defined for $\alpha \in (0,\infty)$ and given by

$$\begin{aligned} h(\alpha) :=& \sum_{y\in\mathcal{Y}} P_Y(y)||P_{X|Y=y}||_\alpha \\ h'(\alpha) =& \sum_{y\in\mathcal{Y}} P_Y(y)||P_{X|Y=y}||_\alpha\Big(\frac{\bar{h}'_y(\alpha)}{\alpha\bar{h}_y(\alpha)} - \frac{\ln(\bar{h}_y(\alpha))}{\alpha^2}\Big) \end{aligned}$$

where for all $y \in \mathcal{Y}$ the function $\bar{h}_y$ and its derivative $\bar{h}'_y$ are continuous and given on $(0, \infty)$ by:

$$\begin{aligned}\bar{h}_y(\alpha) &:= \sum_{x \in \mathcal{X}} P_{X|Y}(x|y)^\alpha \\ \bar{h}'_y(\alpha) &= \sum_{x \in \mathcal{X}} P_{X|Y}(x|y)^\alpha \ln(P_{X|Y}(x|y))\end{aligned}$$

Note, that the computation of the derivatives is easliy done by using the identity $b^\alpha = e^{\alpha \ln(b)}$. Moreover, continuity yields the limits $\lim_{\alpha \to 1} h(\alpha) = 1$ and $\lim_{\alpha \to 1} \bar{h}_y(\alpha) = 1$. In conclusion, by putting everything together we get

$$\begin{aligned}&\lim_{\alpha \to 1} f'(\alpha) = \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{x \in \mathcal{X}} P_{X|Y}(x|y) \log(P_{X|Y}(x|y)) \\ &\lim_{\alpha \to 1} g'(\alpha) = 1\end{aligned}$$

which gives the desired identity with the conditional Shannon entropy:

$$\lim_{\alpha \to 1} H_\alpha(X|Y) = -\sum_{y \in \mathcal{Y}} P_Y(y) \sum_{x \in \mathcal{X}} P_{X|Y}(x|y) \log(P_{X|Y}(x|y))$$

□

The property that will be discussed next is the consistency for the limit of α going to ∞ with the conditional min entropy:

Proposition 4.2. *Let X, Y be random variables. Then,*

$$\lim_{\alpha \to \infty} H_\alpha(X|Y) = H_\infty(X|Y).$$

Proof. For every $\alpha \in (1, \infty)$ one can write as before

$$H_\alpha(X|Y) = -\frac{1}{1 - \frac{1}{\alpha}} \log\Big(\sum_{y \in \mathcal{Y}} P_Y(y) ||P_{X|Y=y}||_\alpha\Big).$$

Now, by $\lim_{\alpha \to \infty} \frac{1}{\alpha} = 0$ and $\lim_{\alpha \to \infty} ||f||_\alpha = ||f||_\infty$ for any function f, it follows that

$$\begin{aligned}\lim_{\alpha \to \infty} H_\alpha(X|Y) &= -\log\Big(\sum_{y \in \mathcal{Y}} P_Y(y) \max_{x \in \mathcal{X}}\{P_{X|Y}(x|y)\}\Big) \\ &= H_\infty(X|Y).\end{aligned}$$

□

Up next is the computation of the limit of α going to 0 yielding a definition for the conditional max entropy:

Proposition 4.3. *Let X, Y be random variables. Then,*

$$H_0(X|Y) := \lim_{\alpha \to 0} H_\alpha(X|Y) = \log\Big(\max_{y \in \mathcal{Y}}\{|\operatorname{supp}(P_{X|Y=y})|\}\Big).$$

Proof. First, define the function $f_\alpha(y) := \sum_{x\in\mathcal{X}} P_{X,Y}(x,y)^\alpha$ on $\mathcal{Y}$ and note that for every $y \in \mathcal{Y}$ it is monotonically increasing for $\alpha \to 0$ with $f_\alpha(y) \to f(y) := |\operatorname{supp} P_{X,Y=y}|$. Next, using this definition of f_α we reformulate the conditional Rényi entropy:

$$H_\alpha(X|Y) = \frac{1}{1-\alpha}\log ||f_\alpha||_{\frac{1}{\alpha}}$$

Now, apply the monotonic increase together with the limit from above:

$$H_\alpha(X|Y) \leq \frac{1}{1-\alpha}\log ||f||_{\frac{1}{\alpha}} \to \log ||f||_\infty \qquad (\alpha \to 0)$$

On the other hand, use $||\cdot||_{\alpha_1} \leq ||\cdot||_{\alpha_2}$ for $\alpha_1 \geq \alpha_2 > 0$:

$$H_\alpha(X|Y) \geq \frac{1}{1-\alpha}\log ||f_\alpha||_\infty \to \log ||f||_\infty \qquad (\alpha \to 0)$$

In conclusion, putting everything together:

$$\lim_{\alpha\to 0} H_\alpha(X|Y) = \log ||f||_\infty = \log\Big(\max_{y\in\mathcal{Y}}\{|\operatorname{supp}(P_{X|Y=y})|\}\Big)$$

□

As a short side note, the subadditivity of the max entropy expands, like one expects, to the conditional case:

Proposition 4.4. *Let X, Y, Z be random variables. The conditional max entropy satisfies subadditivity, i.e.*

$$H_0(X,Y|Z) \leq H_0(X|Z) + H_0(Y|Z).$$

Proof. The obvious fact on the level of set cardinalities

$$\begin{aligned}\max_{z\in\mathcal{Z}}\{|\operatorname{supp}(P_{X,Y|Z=z})\}| &\leq \max_{z\in\mathcal{Z}}\{|\operatorname{supp}(P_{X|Z=z})|\cdot|\operatorname{supp}(P_{Y|Z=z})|\} \\ &\leq \max_{z\in\mathcal{Z}}\{|\operatorname{supp}(P_{X|Z=z})|\}\max_{z\in\mathcal{Z}}\{|\operatorname{supp}(P_{Y|Z=z})|\}\end{aligned}$$

immediately yields the claim on the level of entropies. □

Finally, taking α equal to 2 in our definition yields the following definition for the conditional collision entropy:

Remark. Let X, Y be random variables. Then,

$$H_2(X|Y) = -\log(\operatorname{Col}(X|Y))$$

where

$$\operatorname{Col}(X|Y) := \operatorname{Ren}_2(X|Y) = \Big(\sum_{y\in\mathcal{Y}} P_Y(y)\sqrt{\operatorname{Col}(X|Y=y)}\Big)^2.$$

An immediate consequence is the privacy amplification in case of the conditional collision entropy:

Proposition 4.5. *Let X, Y, S be random variables such that S is independent of (X, Y) and where it is uniformly distributed over $\mathcal{S}$. Let further $g\colon \mathcal{X}\times\mathcal{S} \to \{0,1\}^r$ be a universal hash function, where $1 \leq r < \infty$, and define $K := g(X, S)$. Then,*

$$\Delta[P_{K,Y,S}; U_{\{0,1\}^r} \cdot P_{Y,S}] \leq \frac{1}{2} 2^{-\frac{1}{2}(H_2(X|Y)-r)}$$

Proof. A straightforward computation using Theorem 3.5 yields:

$$\begin{aligned}
\Delta[P_{K,Y,S}; U_{\{0,1\}^r} \cdot P_{Y,S}] &= \sum_{y\in\mathcal{Y}} P_Y(y) \Delta[P_{K,Y,S|Y=y}; U_{\{0,1\}^r} \cdot P_{Y,S|Y=y}] \\
&\leq \sum_{y\in\mathcal{Y}} P_Y(y) \frac{1}{2} 2^{-\frac{1}{2}(H_2(X|Y=y)-r)} \\
&= \frac{1}{2} 2^{\frac{1}{2}r} \sum_{y\in\mathcal{Y}} P_Y(y) \sqrt{\mathrm{Col}(X|Y=y)} \\
&= \frac{1}{2} 2^{\frac{1}{2}r} \sqrt{\mathrm{Col}(X|Y)} \\
&= \frac{1}{2} 2^{\frac{1}{2}r} 2^{-\frac{1}{2}H_2(X|Y)}
\end{aligned}$$

□

4.3. **Properties.** In this subsection, we show that the (new) definition of the conditional Rényi entropy satisfies the properties that we expect from the right notion of conditional Rényi entropy. In particular, we show that monotonicity and (weak) chain rule are satisfied.

4.3.1. *Relation.* In the unconditional case, Proposition 3.8 shows that the Rényi entropy is monotonically decreasing in the parameter α. A general version of this also holds in the conditional case:

Proposition 4.6. *Let X, Y be random variables. If $\alpha, \beta \in [0, \infty]$ with $\alpha \geq \beta$, then $0 \leq H_\alpha(X|Y) \leq H_\beta(X|Y) \leq \log(|\mathcal{X}|)$.*

Proof. The first and last inequality are clear from the definitions. In addition, it is enough to consider $\alpha, \beta \in (0, 1) \cup (1, \infty)$, because otherwise one just takes the corresponding limits. Now, let $\alpha, \beta \in (1, \infty)$, then the claim is equivalent to

$$\Big(\sum_{y\in\mathcal{Y}} P_Y(y) ||P_{X|Y=y}||_\alpha\Big)^{\frac{\alpha}{\alpha-1}} \overset{!}{\geq} \Big(\sum_{y\in\mathcal{Y}} P_Y(y) ||P_{X|Y=y}||_\beta\Big)^{\frac{\beta}{\beta-1}}.$$

Using the fact that $\alpha \geq \beta$, implying $\frac{\alpha(\beta-1)}{(\alpha-1)\beta} \leq 1$, yields in combination with Jensen's inequality at the first inequality and additionally the arguments found in the proof of Proposition 3.8 at the second one:

$$\begin{aligned}\Big(\sum_{y\in\mathcal{Y}} P_Y(y)||P_{X|Y=y}||_\alpha\Big)^{\frac{\alpha}{\alpha-1}} &= \Big(\sum_{y\in\mathcal{Y}} P_Y(y)||P_{X|Y=y}||_\alpha\Big)^{\frac{\alpha(\beta-1)\beta}{(\alpha-1)\beta(\beta-1)}}\\ &\geq \Big(\sum_{y\in\mathcal{Y}} P_Y(y)||P_{X|Y=y}||_\alpha^{\frac{\alpha(\beta-1)}{(\alpha-1)\beta}}\Big)^{\frac{\beta}{\beta-1}}\\ &\geq \Big(\sum_{y\in\mathcal{Y}} P_Y(y)||P_{X|Y=y}||_\beta\Big)^{\frac{\beta}{\beta-1}}\end{aligned}$$

This finishes the case $\alpha, \beta \in (1, \infty)$. On the other hand, the case of $\alpha, \beta \in (0, 1)$ follows by similar arguments. Finally, the case where $\alpha \in (1, \infty)$ and $\beta \in (0, 1)$ follows then by transitivity. □

An immediate consequence of the previous proposition is the equivalent for the conditional case of Proposition 3.6:

Corollary 4.7. *If X, Y are random variables, then*

$$0 \leq H_\infty(X|Y) \leq H_2(X|Y) \leq H_1(X|Y) \leq H_0(X|Y) \leq \log(|\mathcal{X}|).$$

Before moving on to show monotonicity and (weak) chain rule, we state the following observation in preparation:

Proposition 4.8. *Let $\alpha \in [0, \infty]$, $k \in \mathbb{R}$ and X_1, X_2, Y_1, Y_2, Z random variables. If the following statement is true*

$$H_\alpha(X_1|Y_1, Z = z) \geq H_\alpha(X_2|Y_2, Z = z) + k$$

for all $z \in \mathcal{Z}$, then the following statement is true

$$H_\alpha(X_1|Y_1, Z) \geq H_\alpha(X_2|Y_2, Z) + k.$$

Proof. Let us consider $\alpha \in (0, 1) \cup (1, \infty)$, prove the claim in that case and use limits to extend to the cases $\alpha = 0, 1, \infty$. Thus, it is enough to prove the following inequality for $\alpha \in (0, 1) \cup (1, \infty)$:

$$\mathrm{Ren}_\alpha(X_1|Y_1, Z) \overset{!}{\leq} \mathrm{Ren}_\alpha(X_2|Y_2, Z) \cdot 2^{-k}$$

Start with the case $\alpha \in (1, \infty)$. Note that the case $\alpha \in (0, 1)$ follows in the same way, only with reversed inequality at the '$*$'-mark. Now, by assumption the following inequality holds for every $z \in \mathcal{Z}$

$$\mathrm{Ren}_\alpha(X_1|Y_1, Z = z) \leq \mathrm{Ren}_\alpha(X_2|Y_2, Z = z) \cdot 2^{-k}$$

and thus, by taking everything on both sides to the $\frac{\alpha-1}{\alpha}$-th power,

$$\mathrm{Ren}_\alpha(X_1|Y_1, Z = z)^{\frac{\alpha-1}{\alpha}} \overset{*}{\leq} \mathrm{Ren}_\alpha(X_2|Y_2, Z = z)^{\frac{\alpha-1}{\alpha}} \cdot 2^{\frac{(-k)(\alpha-1)}{\alpha}}.$$

Using the inequalities from above together with the following relation

$$\mathrm{Ren}_\alpha(X_i|Y_i, Z) = \Big(\sum_{z\in\mathcal{Z}} P_Z(z)\,\mathrm{Ren}_\alpha(X_i|Y_i, Z = z)^{\frac{\alpha-1}{\alpha}}\Big)^{\frac{\alpha}{\alpha-1}}$$

yields the desired inequality. Note that in the case $\alpha \in (0,1)$ the inequality is reversed once more (negating the other reversion). □

4.3.2. *Monotonicity.*

Proposition 4.9. *Let X, Y be random variables. Then, monotonicity holds for all $\alpha \in [0, \infty]$:*

$$H_\alpha(X) \geq H_\alpha(X|Y)$$

Proof. Let us consider $\alpha \in (0,1) \cup (1, \infty)$, prove the claim in that case and use limits to extend to the cases $\alpha = 0, 1, \infty$. Thus, it is enough to prove the following inequality for $\alpha \in (0,1) \cup (1, \infty)$:

$$\mathrm{Ren}_\alpha(X) \overset{!}{\leq} \mathrm{Ren}_\alpha(X|Y)$$

First, in case of $\alpha \in (1, \infty)$, we simply use the triangle inequality of the α-norm:

$$\begin{aligned}\mathrm{Ren}_\alpha(X) &= \|P_X\|_\alpha^{\frac{\alpha}{\alpha-1}} \\ &= \Big|\Big| \sum_{y \in \mathcal{Y}} P_{X,Y=y} \Big|\Big|_\alpha^{\frac{\alpha}{\alpha-1}} \\ &\leq \Big(\sum_{y \in \mathcal{Y}} ||P_{X,Y=y}||_\alpha \Big)^{\frac{\alpha}{\alpha-1}} \\ &= \Big(\sum_{y \in \mathcal{Y}} P_Y(y) \|P_{X|Y=y}\|_\alpha \Big)^{\frac{\alpha}{\alpha-1}} \\ &= \mathrm{Ren}_\alpha(X|Y)\end{aligned}$$

Second, consider $\alpha \in (0,1)$. By writing the Rényi entropy in terms of the $\frac{1}{\alpha}$-norm, as was derived in the proof of Proposition 4.2, and using the triangle inequality for the $\frac{1}{\alpha}$-norm, we obtain:

$$\begin{aligned}\mathrm{Ren}_\alpha(X) &= \Big(\sum_{x \in \mathcal{X}} ||P^\alpha_{X=x,Y}||_{\frac{1}{\alpha}} \Big)^{\frac{1}{\alpha-1}} \\ &\leq \Big(\Big|\Big| \sum_{x \in \mathcal{X}} P^\alpha_{X=x,Y} \Big|\Big|_{\frac{1}{\alpha}} \Big)^{\frac{1}{\alpha-1}} \\ &= \mathrm{Ren}_\alpha(X|Y)\end{aligned}$$

Note, that in this case, $\frac{1}{\alpha-1} < 0$; Its power reverses inequalities. □

As one might expect, it is possible to extend monotonicity to the case of side information, i.e. conditioning on another random variable:

Corollary 4.10. *Let X, Y be random variables. Then, monotonicity holds conditioned on a random variable Z for all $\alpha \in [0, \infty]$:*

$$H_\alpha(X|Z) \geq H_\alpha(X|Y, Z)$$

Proof. Use Proposition 4.8 with $X_1 = X_2 = X$, $Y_2 = Y$, empty Y_1 and $k = 0$. In case of $\alpha = 0, 1, \infty$, one uses limits again. □

Note that, the proofs of Proposition 4.9 and Corollary 4.10 do not depend on the fact that one is working with probabilty distributions, i.e. everything also applies to non-normalized distributions. Therefore, using the two definitions that follow, we obtain the following:

Corollary 4.11. *Let X, Y, Z be random variables and E an event. Then, monotonicity holds for all $\alpha \in [0, \infty]$:*

$$H_\alpha(X, E|Z) \geq H_\alpha(X, E|Y, Z)$$

In order to undertand the result from above, we need the notion of the Rényi probability of a random variable X with an event E occuring:

Definition 20. *Let X be a random variable and E an event. Then, for $\alpha \in (0, 1) \cup (1, \infty)$,*

$$\mathrm{Ren}_\alpha(X, E) := \Big(\sum_{x \in \mathcal{X}} P_{X,E}(x)^\alpha\Big)^{\frac{1}{\alpha-1}}.$$

The cases $\alpha = 0, 1, \infty$ are defined via limits as usual.

Similarly to this notion, consider the (conditional) Rényi probability of a random variable X and an event E given a random variable Y, yielding the (conditional) Rényi entropy for that case:

Definition 21. *Let X, Y be random variables and E an event. Then, for $\alpha \in (0, 1) \cup (1, \infty)$,*

$$\mathrm{Ren}_\alpha(X, E|Y) := \Big(\sum_{y \in \mathcal{Y}} P_Y(y)(\mathrm{Ren}_\alpha(X, E|Y = y))^{\frac{\alpha-1}{\alpha}}\Big)^{\frac{\alpha}{\alpha-1}}$$
$$H_\alpha(X, E|Y) := -\log(\mathrm{Ren}_\alpha(X, E|Y).$$

The cases $\alpha = 0, 1, \infty$ are defined via limits as usual.

We will use Corollary 4.11 later on.

4.3.3. *Chain rule.*

Proposition 4.12. *Let X, Y be random variables. Then, the chain rule holds for all $\alpha \in [0, \infty]$:*

$$H_\alpha(X|Y) \geq H_\alpha(X, Y) - H_0(Y)$$

Proof. First, note that one only needs to consider $\alpha \in (0, 1) \cup (1, \infty)$ as the claim for $\alpha \in \{0, 1, \infty\}$ follows from taking corresponding limits. Furthermore, consider only $\alpha \in (1, \infty)$ as $\alpha \in (0, 1)$ will follow by similar arguments. The claim is now equivalent to

$$\mathrm{Ren}_\alpha(X|Y) \overset{!}{\leq} \mathrm{Ren}_\alpha(X, Y) \cdot |\mathcal{Y}|$$

Using Jensen's inequality one finds:

$$\begin{aligned}\mathrm{Ren}_\alpha(X|Y) &= \Big(\sum_{y\in\mathcal{Y}} P_Y(y)\|P_{X|Y=y}\|_\alpha\Big)^{\frac{\alpha}{\alpha-1}}\\ &= |\mathcal{Y}|^{\frac{\alpha}{\alpha-1}}\Bigg(\sum_{y\in\mathcal{Y}} \frac{1}{|\mathcal{Y}|}||P_{X,Y=y}||_\alpha\Bigg)^{\frac{\alpha}{\alpha-1}}\\ &\leq |\mathcal{Y}|^{\frac{\alpha}{\alpha-1}}\Bigg(\sum_{y\in\mathcal{Y}} \frac{1}{|\mathcal{Y}|}||P_{X,Y=y}||_\alpha^\alpha\Bigg)^{\frac{1}{\alpha-1}}\\ &= \mathrm{Ren}_\alpha(X,Y)\cdot|\mathcal{Y}|\end{aligned}$$

□

Similar to monotonicity, it is possible to extend the chain rule to the case of side information, i.e. conditioning on another random variable:

Corollary 4.13. *Let X, Y, Z be random variables. Then, the chain rule holds after conditioning on Z for all $\alpha \in [0,\infty]$:*

$$H_\alpha(X|Y,Z) \geq H_\alpha(X,Y|Z) - H_0(Y|Z)$$

Proof. Use $H_0(Y|Z=z) \leq H_0(Y|Z)$ for all $z \in \mathcal{Z}$ and Proposition 4.8 with $X_1 = X$, $X_2 = (X,Y)$, $Y_1 = Y$, empty Y_2 and $k = -H_0(Y|Z)$. Plus, in case of $\alpha = 0, 1, \infty$, one uses limits again. □

Using the previous Corollary repeatedly yields:

Corollary 4.14. *Let $X_1, \ldots, X_n, Z$ be random variables. A generalized chain rule holds for all $\alpha \in [0,\infty]$:*

$$H_\alpha(X_n|(X_j)_{j=1}^{n-1}, Z) \geq H_\alpha((X_j)_{j=1}^n|Z) - \sum_{i=1}^{n-1} H_0(X_i|(X_j)_{j=1}^{i-1}, Z)$$

4.4. **Rényi divergence.** Another aspect of the (new) definition of the conditional Rényi entropy is its relation to the Rényi divergence. Recall the definition:

Definition 22. *Let P, Q be probability distributions over $\mathcal{Z}$. Then, the Rényi divergence D_α of P from Q of order $\alpha \in [0,1)\cup(1,\infty)$ is defined as*

$$D_\alpha(P||Q) := \frac{1}{\alpha-1}\log\Big(\sum_{z\in\mathcal{Z}} P(z)^\alpha Q(z)^{1-\alpha}\Big)$$

and can be extended to the case $\alpha = 1, \infty$ by taking limits:

$$\begin{aligned}D_1(P||Q) &:= \lim_{\alpha\to1} D_\alpha(P||Q)\\ D_\infty(P||Q) &:= \lim_{\alpha\to\infty} D_\alpha(P||Q)\end{aligned}$$

The special cases defined via limits above satisfy the following:

Proposition 4.15. *Let P, Q be probability distributions over $\mathcal{Z}$. Then,*

$$D_1(P||Q) = \sum_{z\in\mathcal{Z}} P(z)\log\left(\frac{P(z)}{Q(z)}\right);$$
$$D_\infty(P||Q) = \log\left(\sup_{z\in\mathcal{Z}}\left\{\frac{P(z)}{Q(z)}\right\}\right).$$

Proof. See for example the original work, [10], for the first equality. □

Remark. In case $\alpha = 1$, the Rényi divergence of P from Q is the Kullback-Leibler divergence (of P from Q).

As indicated before, the Rényi entropy actually bears relation to the Rényi divergence. Let us state it in the unconditional case first:

Remark. Let X be a random variable and furthermore $1_{\mathcal{X}}$ the non-normalized distribution over $\mathcal{X}$ with $1_{\mathcal{X}}(x) = 1$ for all $x \in \mathcal{X}$. Then, for $\alpha \in [0,\infty]$ we find

$$H_\alpha(X) = -D_\alpha(P_X||1_{\mathcal{X}}).$$

Rewriting $1_{\mathcal{X}}$ as $\frac{|\mathcal{X}|}{U_{\mathcal{X}}}$ with $U_{\mathcal{X}}$ the uniform distribution over $\mathcal{X}$ yields

$$H_\alpha(X) = \log|\mathcal{X}| - D_\alpha(P_X||U_{\mathcal{X}}).$$

In other words, the Rényi entropy of X can be understood as the maximal entropy obtained by a uniform and independent $\mathcal{X}$-valued random variable minus how far away the given distribution is from such an 'ideal' one. Something similar holds in the conditional case:

Proposition 4.16. *Let X, Y be random variables with joint probability distribution $P_{X,Y}$ and let the Q_Y's be probability distributions over $\mathcal{Y}$, then for all $\alpha \in [0,\infty]$ it holds:*

$$H_\alpha(X|Y) = \log(|\mathcal{X}|) - \min_{Q_Y} D_\alpha(P_{X,Y}|U_{\mathcal{X}}\cdot Q_Y)$$

Proof. First of all, it is, as usual, enough to consider $\alpha \in (0,1)\cup(1,\infty)$ by using the corresponding limits. Next, let us consider the case $\alpha > 1$ first. A straightforward calculation (using the logarithmic identities with respect to powers and quotients in reverse) shows that the terms on the right hand side are actually equal to the expression

$$-\log\left(\min_{Q_Y}\left\{\sum_{x\in\mathcal{X},y\in\mathcal{Y}} P_{X,Y}(x,y)^\alpha Q_Y(y)^{1-\alpha}\right\}\right)^{\frac{1}{\alpha-1}}.$$

It will be shown below and in detail that $Q_Y(y) = \frac{||P_{X,Y=y}||_\alpha}{\sum_{y'\in\mathcal{Y}}||P_{X,Y=y'}||_\alpha}$ is the minimizing choice. Assuming this result for now, we therefore have

$$\min_{Q_Y}\left\{\sum_{x\in\mathcal{X},y\in\mathcal{Y}} P_{X,Y}(x,y)^\alpha Q_Y(y)^{1-\alpha}\right\} = \left(\sum_{y\in\mathcal{Y}}||P_{X,Y=y}||_\alpha\right)^\alpha$$

which implies that the right hand side is then of the desired form, i.e.

$$-\log\Big(\sum_{y\in\mathcal{Y}} P_Y(y)||P_{X|Y=y}||_\alpha\Big)^{\frac{\alpha}{\alpha-1}} = H_\alpha(X|Y).$$

Thus, only left to show is the claim for the minimizing choice of Q_Y.

Now, let n be the cardinality of Y. In the following, we will consider a fixed enumeration on the n elements of Y. As such one can and will identify (not necessarily normalized) functions $Q_Y\colon Y \to \mathbb{R}_{\geq 0}$ with vectors in $\mathbb{R}^n_{\geq 0}$ and vice versa. Left to do is the optimization using Lagrange multipliers. Define the following functions, $\mathbb{R}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$:

$$\begin{aligned} f(Q_Y) &:= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P_{X,Y}(x,y)^\alpha Q_Y(y)^{1-\alpha} \\ g(Q_Y) &:= \sum_{y\in\mathcal{Y}} Q_Y(y) \end{aligned}$$

Our goal is to maximize $-f(Q_Y)$, i.e. minimize $f(Q_Y)$, under the constraint that $g(Q_Y) = 1$ (the restriction to probability distributions). Therefore, one is looking at the following Lagrange function

$$L(Q_Y, \lambda) : = -f(Q_Y) + \lambda(g(Q_Y) - 1)$$

with the Lagrange multiplier λ.

The partial derivative of L with respect to the unknown $Q_Y(y)$ yield the following expression when set to 0 and simplified:

$$\begin{aligned} 0 &= -\sum_{x\in\mathcal{X}} P_{X,Y}(x,y)^\alpha(1-\alpha)Q_Y(y)^{-\alpha} + \lambda \\ &= -(1-\alpha)Q_Y(y)^{-\alpha}\sum_{x\in\mathcal{X}} P_{X,Y}(x,y)^\alpha + \lambda \end{aligned}$$

In conclusion, $Q_Y(y) = (\frac{1-\alpha}{\lambda})^{\frac{1}{\alpha}}||P_{X,Y=y}||_\alpha$. Using the constraint we obtain $\lambda = (1-\alpha)(\sum_{y'\in\mathcal{Y}}||P_{X,Y=y'}||_\alpha)^\alpha$ and thus the claimed $Q_Y(y)$.

Finally, one needs to check that this is actually a minima. First, the minima can not occur as a point on the boundary because of two observations. Namely, if $Q_Y(y) \to 0$ for $y \in \mathcal{Y}$, then $f(Q_Y) \to \infty$ (implying $Q_Y(y) > 0$). Additionally, we have the constraint $g(Q_Y) = 1$ (implying $Q_Y(y) < \infty$). However, above we have shown that there is only one critical point, which therefore must be a minima.

Now, let us consider the case $\alpha < 1$. Thus, the right hand side is

$$-\log\Big(\max_{Q_Y}\Big\{\sum_{x\in\mathcal{X},y\in\mathcal{Y}} P_{X,Y}(x,y)^\alpha Q_Y(y)^{1-\alpha}\Big\}\Big)^{\frac{1}{\alpha-1}}.$$

Assuming the same Q_Y as before yields again the claimed equality. Thus, similar to before it suffices to analyze its maximizing choice. Modifying the Lagrange optimization from above yields indeed the same possible solution. Using the constraint again (i.e. $Q_Y(y) < \infty$) and the fact that all the partial derivatives of L with respect to $Q_Y(y)$ in this case are actually positive (i.e. $Q_Y(y) > 0$) yields that there is only an interior solution. Thus, the critical point is in fact a maxima. □

Thus, the conditional Rényi entropy of X given Y can therefore be seen as the maximal entropy obtained by a uniform and independent $\mathcal{X}$-valued random variable minus the minimum over distributions on $\mathcal{Y}$ of how far the given joint distribution is away from the product of such a 'varying' one on $\mathcal{Y}$ and an 'ideal' one on $\mathcal{X}$.

4.5. **Splitting.** A useful tool, which is first analyzed with respect to our definition in this subsection and then used later on, is the so called (entropy) splitting. An application of this can be seen in Section 5.

Considering two random variables, splitting yields an lower bound on the entropy of a choice of those two via a third random variable given that very random variable:

Proposition 4.17. *Let $\alpha \in (1, \infty)$ and let X_0, X_1 be random variables. Then, there exists a $\{0, 1\}$-valued random variable C satisfying*

$$H_\alpha(X_{1-C}|C) \geq \frac{H_\alpha(X_0, X_1)}{2} - \frac{\alpha}{\alpha - 1}.$$

Similarly, in case of $\alpha = \infty$, there exists a random variable C with

$$H_\infty(X_{1-C}|C) \geq \frac{H_\infty(X_0, X_1)}{2} - 1.$$

Proof. First, consider the case $\alpha \in (1, \infty)$ and prove $\alpha = \infty$ separately. Using the chain rule proven before and $H_0(C) \leq 1$, it suffices to show there exists a C with:

$$H_\alpha(X_{1-C}, C) \overset{!}{\geq} \frac{d}{2} - \frac{1}{\alpha - 1}$$

where $d := H_\alpha(X_0, X_1)$.

Define C as a function of the random variable X_1 as follows:

$$C := \begin{cases} 0 & P_{X_1}(X_1) < 2^{-\frac{d}{2}} \\ 1 & P_{X_1}(X_1) \geq 2^{-\frac{d}{2}} \end{cases}$$

In conclusion, it follows that

$$\begin{aligned} \operatorname{Ren}_\alpha(X_1, C = 0)^{\alpha-1} &= \sum_{x \in \mathcal{X}} P_{X_1,C}(x, 0)^\alpha \\ &= \sum_{x \in \mathcal{X}} P_{X_1}(x) P_{X_1}(x)^{\alpha-1} P_{C|X_1}(0|x)^\alpha \\ &\leq \left(2^{-\frac{d}{2}}\right)^{\alpha-1} \sum_{x \in \mathcal{X}, P_{X_1}(x) < 2^{-\frac{d}{2}}} P_{X_1}(x) \\ &\leq \left(2^{-\frac{d}{2}}\right)^{\alpha-1} \end{aligned}$$

where the first inequality is due to the fact that for $x \in \mathcal{X}$ it holds that either $P_{X_1}(x) < 2^{-\frac{d}{2}}$ or $P_{X_1}(x) \geq 2^{-\frac{d}{2}}$ and thus by definition of C it holds $P_{C|X_1}(0|x) = 0$. The second inequality follows from the fact that the sum is simply a probability.

On the other hand, by first using monotonicity as in Corollary 4.11

$$\begin{aligned}
\mathrm{Ren}_\alpha(X_0, C=1)^{\alpha-1} &\leq \mathrm{Ren}_\alpha(X_1, C=1|X_1)^{\alpha-1} \\
&= \Big(\sum_{x_1 \in \mathcal{X}_1} P_{X_1}(x_1) || P_{X_0, C=1|X_1=x_1} ||_\alpha \Big)^\alpha \\
&\leq \sum_{x_1 \in \mathcal{X}_1} P_{X_1}(x_1) \sum_{x_0 \in \mathcal{X}_0} P_{X_0,C|X_1}(x_0, 1|x_1)^\alpha \\
&\leq \big(2^{-\frac{d}{2}}\big)^{1-\alpha} \sum_{x_1 \in \mathcal{X}_1} P_{X_1}(x)^\alpha \sum_{x_0 \in \mathcal{X}_0} P_{X_0,C|X_1}(x_0, 1|x_1)^\alpha \\
&\leq \big(2^{-\frac{d}{2}}\big)^{1-\alpha} \sum_{\substack{x_0 \in \mathcal{X}_0 \\ x_1 \in \mathcal{X}_1}} P_{X_0,X_1}(x_0, x_1)^\alpha \\
&\leq \big(2^{-\frac{d}{2}}\big)^{\alpha-1}
\end{aligned}$$

and where the inequalities follow respectively by Jensen's inequality, the definition of C, the fact that additional events reduce probabilities, and the definition of d.

All together this yields:

$$\begin{aligned}
H_\alpha(X_{1-C}, C) &= -\log \Big(\sum_{c \in \{0,1\}} \mathrm{Ren}_\alpha(X_1, C=c)^{\alpha-1} \Big)^{\frac{1}{\alpha-1}} \\
&\geq -\log \left(2 \cdot \left(2^{-\frac{d}{2}}\right)^{\alpha-1} \right)^{\frac{1}{\alpha-1}} \\
&= \frac{d}{2} - \frac{1}{\alpha-1}
\end{aligned}$$

Finally, the case $\alpha = \infty$ follows by taking corresponding limits. □

Extend this now to the case where we additionally condition on a random variable Z (on both sides of the inequality):

Corollary 4.18. *Let $\alpha \in (1, \infty)$ and let X_0, X_1, Z be random variables. Then, there exists a $\{0,1\}$-valued random variable C satisfying*

$$H_\alpha(X_{1-C}|C, Z) \geq \frac{H_\alpha(X_0, X_1|Z)}{2} - \frac{\alpha}{\alpha-1}.$$

Similarly, in case of $\alpha = \infty$, there exists a random variable C with

$$H_\infty(X_{1-C}|C, Z) \geq \frac{H_\infty(X_0, X_1|Z)}{2} - 1.$$

Proof. Use the same approach as the one used for Proposition 4.8. □

At the moment it is not clear, whether this holds for $\alpha \in (0,1)$. However, one can prove the claim for the special case $\alpha = 0$:

Proposition 4.19. *Let X_0 and X_1 be random variables. Then, there exists a $\{0,1\}$-valued random variable C satisfying:*

$$H_0(X_{1-C}|C) \geq \frac{H_0(X_0, X_1)}{2}$$

Proof. In this proof one will use a deterministic C, i.e. $H_0(C) = 0$. Using the chain rule proven before, it suffices to show:

$$H_0(X_{1-C}, C) \overset{!}{\geq} \frac{H_0(X_0, X_1)}{2}$$

Recall Proposition 3.4 about the subadditivity of the max entropy:

$$H_0(X_0) + H_0(X_1) \geq H_0(X_0, X_1)$$

In conclusion, for at least one index $c \in \{0,1\}$ the following holds:

$$H_0(X_c) \geq \frac{H_0(X_0, X_1)}{2}$$

Define $C := 1 - c$ with c such that this inequality is true. Thus,

$$H_0(X_{1-C}, C) = H_0(X_c) \geq \frac{H_0(X_0, X_1)}{2}$$

where the first equality is true as $C = 1 - c$ is deterministic.

□

Of course - as usual - this can also be extended:

Corollary 4.20. *Let X_0, X_1 and Z be random variables. Then, there exists a $\{0,1\}$-valued random variable C satisfying:*

$$H_0(X_{1-C}|C, Z) \geq \frac{H_0(X_0, X_1|Z)}{2}$$

Proof. Use the same approach as the one used for Proposition 4.8. □

The concept or rather property of (entropy) splitting can be extended to the case of more than two random variables. However, similar to the case of two random variables, the bound is probably not optimal, especially considering the limit of α going to 1.

Proposition 4.21. *Let $\alpha \in (1, \infty)$ and let $X_1, \ldots, X_m, Z$ be random variables such that $H_\alpha(X_i, X_j|Z) \geq d$ for all $i \neq j$. Then, there exists a $\{1, \ldots, m\}$-valued random variable U such that for all $\{1, \ldots, m\}$-valued random variables V independent of $(U, X_1, \ldots, X_m, Z)$ that also satisfy $P(V \neq U) \geq p$ it holds that*

$$H_\alpha(X_V|U, V, U \neq V, Z) \geq \frac{d}{2} - \frac{\alpha}{\alpha - 1}\log(m) + \frac{\alpha}{\alpha - 1}\log(p).$$

Similarly, in case of $\alpha = \infty$, it holds that

$$H_\infty(X_V|U, V, U \neq V, Z) \geq \frac{d}{2} - \log(m) + \log(p).$$

The requirement on V with respect to U in the form of $P(V \neq U) \geq p$ for a fixed p is in particular satisfied as long as $H_\infty(V)$ is not too small. In fact, $H_\infty(V) \geq k$ implies that $P(V \neq U) \geq 1 - 2^{-k}$.

Proof. First, consider the case $\alpha \in (1, \infty)$ and prove $\alpha = \infty$ separately. Define for $u \in \{1, \dots, m-1\}$ the set L_u as the set of $(x_1, \dots, x_m, z)$, where $P_{X_k|Z}(x_k|z) < 2^{-\frac{d}{2}}$ for $1 \leq k \leq u-1$ and $P_{X_u|Z}(x_u|z) \geq 2^{-\frac{d}{2}}$. Define U to be the index $u \in \{1, \dots, m-1\}$ for which it holds that $(X_1, \dots, X_m, Z) \in L_u$, else define U to be m. Next, analyze terms depending on the values of V, U.

Let $v < u$, then this yields for every $z \in \mathcal{Z}$

$$\begin{aligned}
\mathrm{Ren}_\alpha(X_v, U = u|Z = z) &= \Big(\sum_{x \in \mathcal{X}_v} P_{X_v|Z}(x|z)^\alpha P_{U|X_v,Z}(u|x,z)^\alpha\Big)^{\frac{1}{\alpha-1}} \\
&\leq 2^{-\frac{d}{2}} \Big(\sum_{x \in \mathcal{X}_v} P_{X_v|Z}(x|z)\Big)^{\frac{1}{\alpha-1}} \\
&= 2^{-\frac{d}{2}}
\end{aligned}$$

where the first inequality follows by the choice of U. In particular,

$$\sum_{z \in \mathcal{Z}} P_Z(z)\, \mathrm{Ren}_\alpha(X_v, U = u|Z = z)^{\frac{\alpha-1}{\alpha}} \leq \big(2^{-\frac{d}{2}}\big)^{\frac{\alpha-1}{\alpha}}.$$

Let $v > u$, then for every $z \in \mathcal{Z}$

$$\begin{aligned}
&\mathrm{Ren}_\alpha(X_v, U = u|Z = z) \\
&\leq \mathrm{Ren}_\alpha(X_v, U = u|X_u, Z = z) \\
&= \Big(\sum_{x_u \in \mathcal{X}_u} P_{X_u|Z}(x_u|z) ||P_{X_v, U=u|X_u=x_u, Z=z}||_\alpha\Big)^{\frac{\alpha}{\alpha-1}} \\
&\leq \Big(\sum_{x_u \in \mathcal{X}_u} P_{X_u|Z}(x_u|z) \sum_{x_v \in \mathcal{X}_v} P_{X_v,U|X_u,Z}(x_v, u|x_u, z)^\alpha\Big)^{\frac{1}{\alpha-1}} \\
&\leq 2^{\frac{d}{2}} \left(\sum_{x_u \in \mathcal{X}_u, x_v \in \mathcal{X}_v} P_{X_u,X_v|Z}(x_u, x_v|z)^\alpha\right)^{\frac{1}{\alpha-1}}
\end{aligned}$$

where one uses Corollary 4.11, Jensen's inequality and definition of U. Together with the assumption on the conditional joint entropy we get

$$\begin{aligned}
&\sum_{z\in\mathcal{Z}} P_Z(z)\,\mathrm{Ren}_\alpha(X_v, U=u|Z=z)^{\frac{\alpha-1}{\alpha}}\\
&\leq \left(2^{\frac{d}{2}}\right)^{\frac{\alpha-1}{\alpha}} \sum_{z\in\mathcal{Z}} P_Z(z) \left(\sum_{x_u\in\mathcal{X}_u, x_v\in\mathcal{X}_v} P_{X_u,X_v|Z}(x_u,x_v|z)^\alpha\right)^{\frac{1}{\alpha}}\\
&\leq \left(2^{-\frac{d}{2}}\right)^{\frac{\alpha-1}{\alpha}}.
\end{aligned}$$

So far, this shows that $\mathrm{Ren}_\alpha(X_v, U=u|Z) \leq 2^{-\frac{d}{2}}$ for all $v \neq u$.

Using $P(V\neq U)\geq p$ together with the restriction with respect to the independence on V at the first inequality, as well as the previous computations in the second:

$$\begin{aligned}
&\Big(\sum_{u,v}\sum_{z\in\mathcal{Z}} P_{U,V,Z|U\neq V}(u,v,z)\Big(\sum_{x\in\mathcal{X}_v} P_{X_V|U,V,Z,U\neq V}(x|u,v,z)^\alpha\Big)^{\frac{1}{\alpha}}\Big)^\alpha\\
&= P(U\neq V)^{-\alpha}\Big(\sum_{\substack{u,v\\u\neq v}}\sum_{z\in\mathcal{Z}} P_{U,V,Z}(u,v,z)\Big(\sum_{x\in\mathcal{X}_v} P_{X_V|U,V,Z}(x|u,v,z)^\alpha\Big)^{\frac{1}{\alpha}}\Big)^\alpha\\
&\leq p^{-\alpha}\Big(\sum_v P_V(v)\sum_{\substack{u\\u\neq v}}\sum_{z\in\mathcal{Z}} P_Z(z)\,\mathrm{Ren}_\alpha(X_v,U=u|Z=z)^{\frac{\alpha-1}{\alpha}}\Big)^\alpha\\
&\leq p^{-\alpha}\Big(\sum_v P_V(v) m \big(2^{-\frac{d}{2}}\big)^{\frac{\alpha-1}{\alpha}}\Big)^\alpha\\
&= p^{-\alpha}\big(2^{-\frac{d}{2}}\big)^{\alpha-1} m^\alpha
\end{aligned}$$

In conclusion, it follows that

$$\begin{aligned}
H_\alpha(X_V|U,V,U\neq V,Z) &\geq -\log\big(p^{-\alpha}\big(2^{-\frac{d}{2}}\big)^{\alpha-1}m^\alpha\big)^{\frac{1}{\alpha-1}}\\
&= \frac{d}{2} - \frac{\alpha}{\alpha-1}\log(m) + \frac{\alpha}{\alpha-1}\log(p).
\end{aligned}$$

Finally, the case $\alpha=\infty$ follows by taking corresponding limits. □

And again, it is not clear, whether this holds for $\alpha\in(0,1)$. However, one can prove the claim for the special case $\alpha=0$:

Proposition 4.22. *Let $X_1,\dots,X_m,Z$ be random variables such that $H_0(X_i,X_j|Z)\geq d$ for all $i\neq j$. Then, there exists a $\{1,\dots,m\}$-valued random variable U such that for all $\{1,\dots,m\}$-valued random variables V independent of $(U,X_1,\dots,X_m,Z)$ it holds that*

$$H_0(X_V|U,V,U\neq V,Z)\geq\frac{d}{2}.$$

Proof. In this proof one will use a deterministic U, i.e. $H_0(U) = 0$. Using the chain rule proven before, it suffices to show:

$$H_0(X_V, U|V, U \neq V, Z) \overset{!}{\geq} \frac{d}{2}$$

Recall Proposition 4.4 about the subadditivity of the max entropy, the conditional version:

$$H_0(X_i|Z) + H_0(X_j|Z) \geq H_0(X_i, X_j|Z)$$

In conclusion, for at least one index $v \in \{1, \ldots, m\}$ the following holds:

$$H_0(X_v|Z) \geq \frac{d}{2}$$

Define $U := u \neq v$ with v such that this inequality is true. Thus,

$$H_0(X_V, U|V, U \neq V, Z) = H_0(X_V|V, V \neq u, Z) \geq H_0(X_v|Z) \geq \frac{d}{2}$$

where the first equality is true as $U = u$ is deterministic and further the second inequality is true by only considering V to be $v \neq u$ and using the restriction with respect to the independence on V. □

5. Quantum ID

In this last section, a particular scheme of quantum cryptographic identification will be our point of interest, namely the one from [8]. Our contribution is a simplified and improved analysis using the (new) notion of conditional collision entropy instead of the so-called smooth min-entropy (or rather its conditional version). Note that we assume the reader to be somewhat familiar with quantum information theory and cryptography as well as with [8].

5.1. **Overview.** The focus of [8] and of our contribution is password-based (quantum) identification with the following setup:

Two parties, U(ser) and S(erver), communicate with each other over a classical and a quantum channel, executing an identification protocol. The identification protocol allows U to 'prove' knowledge of a (possibly low-entropy, human-memorable) pre-agreed password w to S.

One central problem that usually arises in these kind of protocols is the case where one of the parties is dishonest. Either a dishonest S^* tries to impersonate S to learn w from U, possibly in order to impersonate U afterwards to prove w to S. Or similarly a dishonest U^* tries to impersonate U to learn w from S, possibly in order to convince S afterwards of the knowledge of w.

In consequence, the security that one aims for is no leakage of any information about the password that goes beyond the elimination of one possible password per execution of the protocol. Assuming this, a sufficiently large set of possible passwords would allow to safely use the scheme repeatedly. The case of guessing the pre-agreed password without any knowledge about it is of course possible, but its occurance can not be avoided. Additionally, as one expects, the scheme should execute succesfully in case of two honest parties.

In the specific protocol of [8], security is proven under the assumption that any dishonest party has limited quantum memory.

5.2. **Preparation.** Before the scheme is presented in detail, we introduce the notion of a stronlgy universal hash function. In addition to the properties of Definition 17, a hash function g needs to satisfy that $g(x, S)$ and $g(x', S)$ are independent and uniformly distributed over $\mathcal{Z}$ for any $x \neq x'$ (both elements of $\mathcal{X}$). Furthermore, we introduce the following notation and convention:

Notation. Let $I = \{i_1, \dots, i_n\}$ be any index set of $n \in \mathbb{N}$ elements and $\mathcal{X}_i \neq \emptyset$ arbitrary non-empty sets for all $i \in I$. Considering an element $x \in \prod_{i \in I} \mathcal{X}_i$ with $x = (x_{i_1}, \dots, x_{i_n})$ and in addition another index set $J \subseteq I$ with $J = \{j_1, \dots, j_m\}$, denote the restriction of x to J by $x_{|J} \in \prod_{j \in J} \mathcal{X}_j$, i.e. $x_{|J} = (x_{j_1}, \dots, x_{j_m})$.

Convention. Let $x \in \{0,1\}^n$ be an arbitrary n-bit for $n \in \mathbb{N}$ and further $f\colon \{0,1\}^m \to \mathcal{X}$ a hash function for $m \in \mathbb{N}$, $m \leq n$, and $\mathcal{X} \neq \emptyset$.

Consider an index set $J \subseteq \{1, \ldots, n\}$ with $|J| < m$. Applying f to $x_{|J}$ is understood as applying f to $x_{|J}$ padded with sufficiently many 0's.

As indicated before, no extensive general introduction to the topic of quantum theory is given here. Thus, at this point we refer to [7] instead and only state some notations used here:

Notation. When refering to a basis, $\theta \in \{+, \times\}$ is used to indicate the computational basis (or sometimes +-basis), in case of $\theta = +$, or the diagonal basis (sometimes ×- or Hadamard-basis), $\theta = \times$. Thus, $\{|0\rangle_\theta, |1\rangle_\theta\}$ refers to the basis vectors of the basis θ with the relation $|0\rangle_\times = (|0\rangle_+ + |1\rangle_+)/\sqrt{2}$ and $|1\rangle_\times = (|0\rangle_+ - |1\rangle_+)/\sqrt{2}$.

If $x = (x_1, \ldots, x_n) \in \{0,1\}^n$ is an n-bit string and furthermore $\theta = (\theta_1, \ldots, \theta_n) \in \{+, \times\}^n$ an n-basis string, then $|x\rangle_\theta$ is understood as $|x_1\rangle_{\theta_1} \otimes \ldots \otimes |x_n\rangle_{\theta_n}$ - an n-qubit state.

5.3. **Scheme.** An actual execution (of the scheme) takes w, the pre-agreed password, as input and outputs ‘accept’ or ‘reject’.

First, let $\mathcal{W} = \{1, \ldots, m\}$ be the set of available passwords. Furthermore, let $\mathfrak{c} \colon \mathcal{W} \to \{+, \times\}^n$ define an arbitrary (binary) code, but with minimal distance d. It is noteworthy that $\mathfrak{c}$ can be chosen such that n is linear in $\log(m)$ or larger, and d is linear in n. Additionally, let f and g be two strongly universal hash functions defined respectively from $\{0,1\}^n \times \mathcal{S}_f$ to $\{0,1\}^l$ and from $\mathcal{W} \times \mathcal{S}_g$ to $\{0,1\}^l$. Then, the actual protocol goes as follows:

(1) U picks uniformly $x \in \{0,1\}^n$ and $\theta \in \{+, \times\}^n$, and sends $|x\rangle_\theta$;
(2) S measures $|x\rangle_\theta$ in basis $\mathfrak{c}(w)$ with outcome x';
(3) U picks uniformly $s_f \in \mathcal{S}_f$ and sends θ and s_f to S - both compute I_w, where $I_w := \{i \,|\, \theta_i = \mathfrak{c}(w)_i\}$;
(4) S picks uniformly $s_g \in \mathcal{S}_g$, and sends s_g to U;
(5) U computes and sends z to S, where $z := f(x_{|I_w}, s_f) \oplus g(w, s_g)$;
(6) S accepts if and only if $z = z'$, where $z' := f(x'_{|I_w}, s_f) \oplus g(w, s_g)$.

In the analysis below, the choice of w is described via the random variable W; the ones for x and θ via X and Θ respectively as well as the ones for s_f and s_g via S_f and S_g.

The assumption about the limited quantum memory is incorporated such that a dishonest party can store at most q qubits before step 3 is executed. In case of a dishonest S^* the resulting q-qubit state of S^* is denoted by E_{S^*}.

5.4. **Analysis.** [8] used the so-called (conditional) smooth min entropy as the measure of uncertainty, while in our security analysis we will use the (conditional) collision entropy. Concretely, the security analysis in [8] is based on two results: An entropic quantum uncertainty relation which lower bounds the conditional smooth min entropy and an entropy splitting result for the conditional smooth min entropy. We point out that the entropy splitting blows up the smoothing parameter unfavorably and results in a rather complicated expression, which we can avoid altogether in our analysis.

In order to do our analysis, we need the corresponding tools for the (conditional) collision entropy. In Section 4 of this thesis we already introduced the entropy splitting and so what remains is a corresponding uncertainty relation:

Proposition 5.1. *Let E be an arbitrary fixed n-qubit state. Further, let the random variable Θ (independent of E) be uniformly distributed over $\{+,\times\}^n$ and let X be the $\{0,1\}^n$-valued random variable for the outcome of measuring E in basis Θ. Then,*

$$H_2(X|\Theta) \geq -\log\left(\frac{3}{4}\right)n \approx 0.415n.$$

Proof. This was proven in unpublished notes (a preliminary version of [8]). Actually, the lower bound was proven for a different version of conditional collision entropy (namely for H_2^2, using the notation from Section 3.4), but it follows immediately from Jensen's inequality that the bound also applies to our notion of conditional collision entropy. □

In the following, our focus will be on the security of the scheme with respect to a dishonest server (i.e. user security). Similar to the analysis of [8] it will be our aim to show that $f(X_W, S_f)$ is (almost) uniformly distributed. However, this asks for too much. More accurately, we will show (almost) uniformity of $f(X_W, S_f)$ conditioned on the event $W \neq W'$, where W' is a random variable that is well defined by S^*'s strategy and is independent of W. This captures that S^* can always make a guess for W, and there is nothing we can hope for in case that his guess is right. For the formal security definition consult [8].

Proposition 5.2. *Q-ID is secure for the user with error ϵ against S^* under the assumption that $H_\infty(W) \geq 1$ and where*

$$\epsilon = 2^{-(\frac{1}{10}d - \log(m) - \frac{1}{2}q - \frac{1}{2}l)}.$$

Proof (sketch). First of all, as in [8], for our analysis we consider a 'purified' version of the scheme. In step 1, U picks θ as before, but prepares $2^{-\frac{n}{2}} \sum_{y \in \{0,1\}^n} |y\rangle|y\rangle$ and sends the second register instead of preparing and sending $|x\rangle_\theta$. Consequently, in step 3, U measures the first register in the basis θ to obtain x. Note, that the purified version still yields the same common state in the end.

Let us define $X_w := X_{I_w}$ for all $w \in \mathcal{W}$. Fix $u, v \in \mathcal{W}$ with $u \neq v$ and define the set $I := \{i \mid \mathfrak{c}(u)_i \neq \mathfrak{c}(v)_i\}$ satisfying $I \subseteq I_u \cup I_v$. In addition, using the minimal distance d of $\mathfrak{c}$ yields $|I| \geq d$. Thus, the inclusion of the index sets together with Proposition 5.1, the uncertainty relation, give:

$$H_2(X_u X_v|\Theta) \geq H_2(X_I|\Theta) \geq \frac{4}{10}d$$

Therefore, by applying Proposition 4.21, the entropy splitting, it holds that there exists a random variable W' independent of W such that

$$H_2(X_W|W'W\Theta, W' \neq W) \geq \frac{2}{10}d - 2\log(m) - 2.$$

Note that the fact $H_\infty(W) \geq 1$, implying $P(W' \neq W) \geq \frac{1}{2}$, is used.

Using an extension of Proposition 4.5, i.e. the privacy amplification theorem, to the quantum case (cf. [9]) yields that

$$\rho_{f(X_W,S_f)S_fWW'\Theta E_{S^*}|W'\neq W} \approx_\epsilon \frac{1}{2^l}\mathbb{I} \otimes \rho_{S_fWW'\Theta E_{S^*}|W'\neq W}$$

with

$$\epsilon := \frac{1}{2}2^{-\frac{1}{2}(\frac{2}{10}d-2\log(m)-2-q-l)}.$$

□

Notice, that the (result of the) analysis of the security of the scheme with respect to a dishonest user (i.e. server security) found in [8] is independent of the smoothing parameter.

Proposition 5.3. *Q-ID is secure for the server with error ϵ' against U^* under the assumption that $H_\infty(W) \geq 1$ and where*

$$\epsilon' = 2^{-(-2\log(m)+l)}.$$

Putting $\epsilon = \epsilon'$ yields $l = \frac{1}{15d} + \frac{2}{3}\log(m) - \frac{1}{3}q$ and therefore:

Proposition 5.4. *If $H_\infty(W) \geq 1$, then Q-ID is ϵ-secure against any dishonest party where*

$$\epsilon = 2^{-(\frac{1}{15}d-\frac{4}{3}\log(m)-\frac{1}{3}q)}.$$

In conclusion, in order to have at most an exponentially small error of 2^{-k} for $k \in \mathbb{N}$ the factor q needs to be upper bounded as follows:

$$q \leq \frac{1}{5}d - 4\log(m) - 3k$$

Finally, for comparison let us state the analog to Proposition 5.4 that is given in [8]:

Proposition 5.5. *If $H_\infty(W) \geq 1$, then Q-ID is $\tilde{\epsilon}$-secure against any dishonest party where*

$$\tilde{\epsilon} = 2^{-((\frac{1}{12}-\frac{1}{3}\lambda)d-\log(m)-\frac{1}{3}q-\frac{2}{3})} + 2^{-(\sigma(\lambda)d-\log(m)-4)}$$

for an arbitrary $0 < \lambda < \frac{1}{4}$ and $\sigma(\lambda) := \frac{\lambda^2\log(e)}{32(2-\log(\lambda))^2}$.

6. Conclusion

The contribution of this thesis is the introduction of a new notion of conditional Rényi entropy. We showed that it satisfies all the properties one would naturally expect from a 'good' entropy notion. Additionally, its natural interpretation with respect to the notion of Rényi divergence supports the claim that it is the 'right' definition. The discussion of its application to the privacy amplification theorem and the particular scheme of quantum cryptographic identification allowed a glimpse at its potential use in the field of (quantum) cryptography.

Potential follow-ups to this thesis could for example aim for (more) applications or uniqueness with respect to all the natural properties.

References

[1] L. Antunes, A. Matos, and A. Teixeira, *Conditional rényi entropies*, IEEE Transactions on Information Theory **58** (2012), no. 7, 4273–4277.

[2] C. Cachin, *Entropy measures and unconditional security in cryptography*, Ph.D. thesis, Swiss Federal Institute of Technology Zürich, 1997.

[3] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, *Fuzzy extractors: How to generate strong keys from biometrics and other noisy data*, SIAM Journal on Computing **38** (2008), no. 1, 97–139.

[4] R. Impagliazzo, L. A. Levin, and M. Luby, *Pseudo-random generation from one-way functions*, Proceedings of the 21st Annual ACM Symposium on Theory of Computing (1989), 12–24.

[5] J. L. W. V. Jensen, *Sur les fonctions convexes et les inégalités entre les valeurs moyennes*, Acta Mathematica **30** (1906), no. 1, 175–193.

[6] H. Kaizhu, I. King, and M.R: Lyu, *Direct zero-norm optimization for feature selection*, Proceedings of the eighth IEEE International Conference on Data Mining (2008), 845–850.

[7] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, 2000.

[8] I. Damgård, S. Fehr, L. Salvail, and C. Schaffner, *Secure identification and qkd in the bounded-quantum-storage model*, Published at Crypto 2007; cf. http://arxiv.org/abs/0708.2557.

[9] R. Renner and R. König, *Universally composable privacy amplification against quantum adversaries*, Theory of Cryptography Conference, 2005, pp. 407–425.

[10] A. Rényi, *On measures of information and entropy*, Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960 (1961), 547–561.

[11] C. E. Shannon, *A mathematical theory of communication*, Bell System Technical Journal **27** (1948), no. 3, 379–423.

[12] B. Škorić, C. Obi, E. A. Verbitskiy, and B. Schoenmakers, *Sharp lower bounds on the extractable randomness from non-uniform sources*, Information and Computation **209** (2011), no. 8, 1184–1196.