

TECHNISCHE UNIVERSITEIT EINDHOVEN
Department of Mathematics and Computer Science

Smooth Rényi Entropy of
Ergodic Quantum Information Sources
by
J. Tjoelker

Supervisors:
Dr.Ir. L.A.M. Schoenmakers, Dr. P. Tuijls

Eindhoven, April 2007

Abstract

We investigate the recently introduced notion of smooth Rényi entropy, comparing the slightly different definitions and studying the case of *ergodic* information sources, thereby generalizing previous work which concentrated mainly on i.i.d. information sources. We will actually consider ergodic *quantum* information sources, of which ergodic *classical* information sources are a special case. We prove that the average smooth Rényi entropy rate will approach the entropy rate of a stationary, ergodic source, which is equal to the Shannon entropy for a classical source and the von Neumann entropy for a quantum source.

Acknowledgements

First of all, I would like to thank Berry Schoenmakers and Pim Tuyls for providing me with a challenging problem and for the helpful suggestions and comments. Further input from Evgeny Verbitskiy and Boris Škorić is also acknowledged.

Henk van Tilborg and Frans Willems, thank you for being in my committee.

Then I want to thank Sandra van Dongen for support during the final years of my studies, mainly in getting the final pieces before the graduation phase done.

Last but not least I thank my parents for giving me the opportunity to do my studies in the first place.

Contents

1	Introduction	3
1.1	Background	3
1.2	Information-theoretically secure key agreement	4
1.3	Rényi entropy and smooth Rényi entropy	4
1.4	Outline of this thesis	5
2	Preliminaries	6
2.1	Probability distributions	6
2.2	Linear algebra	6
2.3	Quantum mechanics	7
2.4	Distance measures	8
2.5	Entropy measures	8
2.6	Ergodicity	12
2.6.1	Classical case	12
2.6.2	Quantum case	13
2.7	Asymptotic equipartition property	14
2.7.1	Classical case	14
2.7.2	Quantum case	15
3	A closer look at smooth Rényi entropy	16
3.1	Comparison of the two definitions, classical case	16
3.1.1	Maps between the balls	16
3.1.2	Bounds	18
3.1.3	Example of the difference	20
3.2	Comparison of the two definitions, quantum case	21
3.2.1	Comparison between quantum and classical information	21
3.2.2	Bounds	24
3.3	Relations between smooth Rényi entropy of different orders	25
4	Asymptotic results for smooth Rényi entropy	27
4.1	Definitions (classical case)	27
4.2	Classical case, truncation ball	27
4.3	Classical case, statistical distance ball	30
4.4	Quantum case, truncation ball, indirect proof	32
4.5	Quantum case, trace distance ball, indirect proof	33
4.6	Quantum case, truncation ball, direct proof	33

5 Conclusions

38

Chapter 1

Introduction

1.1 Background

Nowadays, much information is exchanged digitally. Often this information needs to be protected against unauthorized people reading it (encryption) and unauthorized changes (authentication).

This requires key material, random bits known to the communicating parties but not to others. Key exchange is the process of establishing this key material. Knowing the key material allows normal usage of the system.

Currently, most information protection is “computationally secure”. This means that it is possible, given enough computer time and a sufficiently clever algorithm, to find out the encrypted message or change the authenticated message without knowing the key. The systems are tuned such that, given the current and expected future state of computer technology and algorithm theory, it is easy to use the system normally (i.e. knowing the key) but infeasible to break it.

This thesis is related to “information-theoretically secure” systems, sometimes known as “unconditionally secure” systems. In such systems, the best strategy to find out the encrypted message or to change the authenticated message is guessing the key, no matter how much computer time is used. In other words, the encrypted message does not give an attacker any information at all about the plaintext. More information about this distinction can be found in [9].

Quantum mechanics has several interesting consequences for cryptography. Firstly, algorithms for quantum computers exist that can solve the discrete log problem and prime factorization in polynomial time, so cryptographic systems relying on their intractability are no longer secure. Fortunately, quantum computers can currently only process very small sets of data, and this is unlikely to change in the near future.

Secondly, the fact that the only means to get information from a quantum system is measuring it, which changes the system (i.e. eavesdropping can be detected), enables new information-theoretically secure systems.

1.2 Information-theoretically secure key agreement

Using quantum mechanics or some other “special” channels between the two parties, a class of information-theoretically secure key exchange protocols can be constructed. A second requirement is the existence of a classical channel between the parties, which the attacker can read but not change. An example of such a special channel is a satellite sending out random bits at very low signal power, which nobody can receive 100% correctly ([9, section 5.2]).

These protocols consist of three steps: advantage distillation, information reconciliation and privacy amplification.

In advantage distillation, an advantage is obtained over the attacker. After this step, both parties have a string of bits. The strings of bits of both parties are not necessarily equal and the attacker may have some information about them, but the parties have more information about each other’s strings than the attacker does.

In information reconciliation, the parties agree on a string by exchanging information on the classical channel. This also gives some additional information to the attacker. An interesting question is how much information needs to be exchanged to agree on a string.

In privacy amplification, the string is replaced by a smaller string which the attacker has negligible information about. This is done by negotiating a compression function over the classical channel, and applying the function to the string. This negotiation must not start before the previous steps are done, lest the attacker collect information in such a way that the compression function does not reduce it. An interesting question is how long this smaller string, the key, can safely be.

1.3 Rényi entropy and smooth Rényi entropy

Important concepts in answering the above questions are Rényi entropy and smooth Rényi entropy. They quantify the worst case in information reconciliation (how much information needs to be exchanged) and privacy amplification (how long can the key safely be), where Shannon entropy (the most basic measure of information quantity) would quantify the average case.

More information about privacy amplification (also known as entropy smoothing in a more general context, for example complexity theory) and the difference between Shannon and Rényi entropy can for example be found in [3], which predates the concept of smooth Rényi entropy. It defines “smooth entropy” as the highest number of bits of uniform randomness that can be extracted after redistributing ϵ of probability mass in the best possible way (this corresponds to allowing the operation to fail with probability ϵ). It does not specify how to calculate this quantity exactly, only giving bounds.

Smooth Rényi entropy was introduced by Renner and Wolf in [7] and [8], combining Rényi entropy and the redistribution of ϵ of probability mass. This allows

calculating smooth entropy and related notions. The resulting quantity has much better properties than regular Rényi entropy.

1.4 Outline of this thesis

The goal of this thesis is to prove that smooth Rényi entropy is equal to Shannon or Von Neumann entropy in the limit for the number repetitions goes to infinity and $\epsilon \rightarrow 0$.

We will describe Rényi entropy and smooth Rényi entropy. These quantities can be defined for both classical information and quantum information (section 2.5).

We will describe two variations of smooth Rényi entropy, from the two papers, which differ in the definition of the ball used for smoothing. In either case, the infimum or supremum is taken over a ball around the value, but one definition (statistical distance ball) uses all probability distributions which are close enough, while the other (truncation ball) cuts off some probability mass so that the elements of the ball are not probability distributions.

As a generalization of independent identically distributed repetitions, we will consider stationary ergodic information sources (section 2.6).

Stationary ergodic information sources have an important property: given enough repetitions, most probability mass is in the “typical set” where every possibility has approximately the same probability. As the number of repetitions increases, the probability mass in the typical set increases and the probabilities get closer together. This is called the AEP (asymptotic equipartition property). Our proofs will use the AEP and will not use ergodicity directly.

We will give bounds (for the classical case only) for the difference between the two variations of smooth Rényi entropy (section 3.1 and 3.2).

We will prove that, given enough repetitions, smooth Rényi entropy (both variations) is equal to Shannon entropy in the limit. These repetitions do not have to be independent and identically distributed; it is sufficient if they are stationary and ergodic (chapter 4).

Chapter 2

Preliminaries

2.1 Probability distributions

Definition 2.1 (Probability distribution) *A probability distribution is a function \mathbb{P} from a set \mathcal{Z} to \mathbb{R} such that $\forall_{z \in \mathcal{Z}} \mathbb{P}(z) \geq 0$ and $\sum_{z \in \mathcal{Z}} \mathbb{P}(z) = 1$.*

Except when otherwise noted, the set \mathcal{Z} will be finite.

We will define the entropy measures on a generalization of probability distributions:

Definition 2.2 (Non-normalized probability distribution) *A non-normalized probability distribution is a function \mathbb{P} from a set \mathcal{Z} to \mathbb{R} such that $\forall_{z \in \mathcal{Z}} \mathbb{P}(z) \geq 0$ and $0 < \sum_{z \in \mathcal{Z}} \mathbb{P}(z) \leq 1$.*

2.2 Linear algebra

This section will review a few concepts of linear algebra.

Definition 2.3 (Projection) *A square matrix P is a projection if all eigenvalues are 0 or 1.*

Theorem 2.4 *For a projection P , $\text{tr}(P) = \text{rank}(P)$.*

Definition 2.5 (Hermitian matrix) *A Hermitian matrix is a matrix A such that $A = A^\dagger$, where A^\dagger is the conjugate transpose of A .*

Theorem 2.6 *A matrix is Hermitian if and only if it is diagonalizable, all eigenvalues are real and the eigenvectors are orthogonal.*

Definition 2.7 (Positive matrix) *A positive matrix (sometimes known as a positive semidefinite matrix) is a matrix whose eigenvalues are all nonnegative real.*

We will also use the notation $A \geq 0$ to state that a matrix A is positive, and the notation $A \geq B$ to state that $A - B$ is positive.

Theorem 2.8 *A positive matrix is Hermitian.*

For comparison, we will give the following definition; we will not use it.

Definition 2.9 (Strictly positive matrix) *A strictly positive matrix (sometimes known as a positive definite matrix) is a matrix whose eigenvalues are all positive real.*

Theorem 2.10 *If P is a projection and $A \geq 0$, then $PA \geq 0$ and $AP \geq 0$.*

2.3 Quantum mechanics

Only a terse description will be given here, more detailed information can for example be found in [5], chapter 2. This description will be completely mathematical; the physics will not be discussed.

A quantum state is a representation of the state of a physical system by a vector of length 1 in a Hilbert space. If the Hilbert space is \mathbb{C}^2 the quantum state is called a qubit. An orthonormal basis is $\{|0\rangle, |1\rangle\}$, corresponding to 0 and 1 for classical bits.

The only way to get information out of a quantum system is by performing a measurement. A measurement is described by a collection $\{M_m\}$ of matrices over the state space, where m are the possible measurement outcomes. If the state of the system before the measurement is $|\psi\rangle$, the probability of measurement outcome m is $p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle$. After the measurement the state will “collapse” to $\frac{M_m|\psi\rangle}{p(m)}$ if the outcome is m .

In what follows one party will often send one of several states with certain probabilities, for example $|0\rangle$ with probability 1/2 and $|1\rangle$ with probability 1/2. Due to the fact that the receiver can only distinguish the states by performing a measurement, this can be represented more compactly with a matrix over the state space, called a density matrix or density operator:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$$

In the example the density matrix is

$$\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

The sender could also send $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ with probability 1/2 and $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ with probability 1/2, and the density matrix would be the same.

In the following this characterization will be most useful.

Definition 2.11 (Density matrix) *A density matrix is a positive matrix whose eigenvalues sum to one.*

The following generalization of density matrices will be useful:

Definition 2.12 (Non-normalized density matrix) *A non-normalized density matrix is a positive matrix whose eigenvalues sum to a number t with $0 < t \leq 1$.*

2.4 Distance measures

Measures of distance between probability distributions are needed.

Statistical distance is a distance measure for the classical case.

Definition 2.13 (Statistical distance) *Given two probability distributions \mathbb{P} and \mathbb{Q} over \mathcal{Z} , the statistical distance between them is*

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

Trace distance is the quantum analogon of statistical distance.

Definition 2.14 (Trace distance) *Given two density matrices ρ and σ , the trace distance between them is*

$$\delta(\rho, \sigma) = \frac{1}{2} \text{tr} |\rho - \sigma|$$

where $|A| = \sqrt{A^\dagger A}$.

If ρ and σ commute and the eigenvalues of ρ and σ are $\mathbb{P}(i)$ and $\mathbb{Q}(i)$ respectively, then $\delta(\rho, \sigma) = \delta(\mathbb{P}, \mathbb{Q})$. This property is useful in some proofs.

2.5 Entropy measures

Entropy measures quantify the amount of uncertainty in a probability distribution. The most basic entropy measure is Shannon entropy.

Definition 2.15 (Shannon entropy) *The Shannon entropy of a non-normalized probability distribution \mathbb{P} over \mathcal{Z} is defined as*¹

$$H(\mathbb{P}) = - \sum_{i \in \mathcal{Z}} \mathbb{P}(i) \log \mathbb{P}(i)$$

Von Neumann entropy is the quantum analogon of Shannon entropy.

¹All logarithms in this thesis have base 2.

Definition 2.16 (Von Neumann entropy) *Von Neumann entropy of a non-normalized density matrix is defined as*

$$S(\rho) = -\text{tr}(\rho \log \rho)$$

Let λ_i denote the eigenvalues of the non-normalized density matrix ρ . Then $S(\rho) = -\sum_i \lambda_i \log \lambda_i$.

Rényi entropy is a family of entropy measures. It has a parameter (order) $\alpha \in [0, \infty]$.

Definition 2.17 (Rényi entropy (classical)) *In the classical case, Rényi entropy of order α is defined as*

$$H_\alpha(\mathbb{P}) = \frac{1}{1-\alpha} \log \left(\sum_{z \in \mathcal{Z}} \mathbb{P}(z)^\alpha \right)$$

for a non-normalized probability distribution \mathbb{P} and $0 < \alpha < 1 \vee 1 < \alpha < \infty$, with the convention that $H_\alpha(\mathbb{P}) = \lim_{\beta \rightarrow \alpha} H_\beta(\mathbb{P})$ for $\alpha \in \{0, 1, \infty\}$.

For $\alpha = 0$, this gives $H_0(\mathbb{P}) = \log(\#\{z \in \mathcal{Z} | \mathbb{P}(z) > 0\})$. For $\alpha = 1$, this gives $H_1(\mathbb{P}) = H(\mathbb{P})$ (Shannon entropy). For $\alpha = \infty$, this gives $H_\infty(\mathbb{P}) = -\log(\max\{\mathbb{P}(z) | z \in \mathcal{Z}\})$.

Note that Rényi entropy of all orders is the same as Shannon entropy for a distribution \mathbb{P} with $\mathbb{P}(i) = \frac{1}{n}$ for $i = 1, \dots, n$:

$$H_\alpha(\mathbb{P}) = \frac{1}{1-\alpha} \log \left(n \left(\frac{1}{n} \right)^\alpha \right) = \frac{1}{1-\alpha} (\log n - \alpha \log n) = \log n.$$

Rényi entropy of order 0 is log of the number of elements in \mathcal{Z} with non-zero probability. This is important in information reconciliation and privacy amplification, describing the amount of information needed to reconstruct a value exactly without chance of failure.

For example, consider a random variable X with $\mathbb{P}(X = 1) = \frac{1}{2}$ and $\mathbb{P}(X = i) = \frac{1}{2(n-1)}$ for $i = 2, \dots, n$. Then $H(\mathbb{P}) = \frac{1}{2} + \frac{1}{2} \log 2(n-1) = \frac{1}{2} \log 4(n-1) = \log 2\sqrt{n-1}$. However $H_0(\mathbb{P}) = \log n$, reflecting that n bits are needed to reconstruct X without chance of failure.

Rényi entropy of order 2 is the negative log of the probability that two independent repetitions of $\mathbb{P}(z)$ give the same element of \mathcal{Z} .

Rényi entropy of order ∞ is the negative log of the largest probability. This is important in cryptography, describing the probability of an attacker guessing the key. In the example, $H_\infty(\mathbb{P}) = 1$ and an attacker has 50% probability of guessing X right, much higher than the Shannon entropy suggests.

Note that independent identically distributed repetitions do not remove this gap between Shannon entropy and Rényi entropy.

Definition 2.18 (Rényi entropy (quantum)) *In the quantum case, Rényi entropy is defined as ([6])*

$$S_\alpha(\rho) = \frac{1}{1-\alpha} \log(\text{tr}(\rho^\alpha))$$

for a non-normalized density matrix ρ and $0 < \alpha < 1 \vee 1 < \alpha < \infty$, with the convention that $S_\alpha(\rho) := \lim_{\beta \rightarrow \alpha} S_\beta(\rho)$ for $\alpha \in \{0, 1, \infty\}$.

Equivalently, if $p(z)$ ($z \in \mathcal{Z}$) are the eigenvalues of ρ , $S_\alpha(\rho) = H_\alpha(p(z))$.

Then $S_0(\rho) = \log(\text{rank}(\rho))$, $S_1(\rho) = S(\rho)$ (Von Neumann entropy) and $S_\infty(\rho) = -\log \lambda_{\max}(\rho)$ (maximum eigenvalue).

Smooth Rényi entropy takes the infimum ($\alpha < 1$) or supremum ($\alpha > 1$) of the Rényi entropy over all \mathbb{Q} which are close to \mathbb{P} in some way (parametrized by $\epsilon > 0$). For $\alpha = 1$, smooth Rényi entropy is the same as Shannon entropy.

We will use two different definitions of “close” (represented by two different balls $\mathcal{B}^\epsilon(\mathbb{P})$). The old definition from [7] uses statistical distance:

$$\mathcal{B}_o^\epsilon(\mathbb{P}) := \{\mathbb{Q} \mid \delta(\mathbb{P}, \mathbb{Q}) \leq \epsilon, \sum_z \mathbb{Q}(z) = 1, \forall_z \mathbb{Q}(z) \geq 0\}.$$

In the new definition from [8] (also mentioned in the full version of [7]) the elements of the ball are not probability distributions (except \mathbb{P}); however, they are non-normalized probability distributions. We will call this the truncation ball:

$$\mathcal{B}_n^\epsilon(\mathbb{P}) := \{\mathbb{Q} \mid \sum_{z \in \mathcal{Z}} \mathbb{Q}(z) \geq 1 - \epsilon, \forall_z \mathbb{Q}(z) \leq \mathbb{P}(z), \forall_z \mathbb{Q}(z) \geq 0\}.$$

Note that both balls are compact sets, because they are closed bounded subsets of $\mathbb{R}^{\#\mathcal{Z}}$.

Putting \mathcal{B}_o^ϵ (statistical distance ball) or \mathcal{B}_n^ϵ (truncation ball) for \mathcal{B}^ϵ , the rest of the definition is as follows:

Definition 2.19 (Smooth Rényi entropy (classical))

$$H_\alpha^\epsilon(\mathbb{P}) = \frac{1}{1-\alpha} \inf_{\mathbb{Q} \in \mathcal{B}^\epsilon(\mathbb{P})} \log \left(\sum_{z \in \mathcal{Z}} \mathbb{Q}(z)^\alpha \right), \quad \text{for } 0 < \alpha < 1 \vee 1 < \alpha < \infty$$

where the special cases for 0, 1 and ∞ are the limits as in Rényi entropy:

$$\begin{cases} H_\alpha^\epsilon(\mathbb{P}) = \inf_{\mathbb{Q} \in \mathcal{B}^\epsilon(\mathbb{P})} \log(\#\{z \in \mathcal{Z} \mid \mathbb{Q}(z) > 0\}), & \text{for } \alpha = 0 \\ H_\alpha^\epsilon(\mathbb{P}) = H(\mathbb{P}), & \text{for } \alpha = 1 \\ H_\alpha^\epsilon(\mathbb{P}) = -\inf_{\mathbb{Q} \in \mathcal{B}^\epsilon(\mathbb{P})} \log(\max\{\mathbb{Q}(z) \mid z \in \mathcal{Z}\}), & \text{for } \alpha = \infty \end{cases}$$

Another formulation of the definition, which can be easier to use:

$$H_\alpha^\epsilon(\mathbb{P}) = \begin{cases} \inf_{\mathbb{Q} \in \mathcal{B}^\epsilon(\mathbb{P})} H_\alpha(\mathbb{Q}), & \text{for } \alpha < 1; \\ \sup_{\mathbb{Q} \in \mathcal{B}^\epsilon(\mathbb{P})} H_\alpha(\mathbb{Q}), & \text{for } \alpha > 1. \end{cases}$$

The following theorem is already given by Renner and Wolf in [8, section 2.1]; we will give the (simple) proof.

Theorem 2.20 *The infimum or supremum in the definition of classical smooth Rényi entropy is actually a minimum or maximum.*

Proof Use the second formulation of the definition.

For $\alpha = 0$, the function H_0 maps $\mathbb{R}^{\#\mathcal{Z}}$ to a subset of $\{\log M | M \in \mathbb{N}, 1 \leq M \leq \#\mathcal{Z}\}$ which is a finite set, so it has a minimum and a maximum.

For $\alpha > 0$, H_α is a continuous function from a subset of $\mathbb{R}^{\#\mathcal{Z}}$ to \mathbb{R} , so it maps the compact set $\mathcal{B}^\epsilon(\mathbb{P})$ to a compact subset of \mathbb{R} which has a minimum and a maximum. \square

In the quantum case the equivalent of the ball using statistical distance is a ball using trace distance:

$$\mathcal{B}_0^\epsilon(\rho) := \{\sigma | \delta(\rho, \sigma) \leq \epsilon, \sigma \text{ is a density operator}\}$$

The equivalent of the truncation ball is:

$$\mathcal{B}_n^\epsilon(\rho) := \{\sigma \geq 0 | \sigma \leq \rho, \text{tr}(\sigma) \geq 1 - \epsilon\}$$

Similarly to the classical case, the elements of the truncation ball are not density matrices, but they are non-normalized density matrices.

Both balls are compact sets.

Definition 2.21 (Smooth Rényi entropy (quantum)) *In either case the smooth Rényi entropy is defined as:*

$$S_\alpha^\epsilon(\rho) := \frac{1}{1 - \alpha} \inf_{\sigma \in \mathcal{B}^\epsilon(\rho)} \log(\text{tr}(\rho^\alpha))$$

with the convention that $S_\alpha^\epsilon(\rho) := \lim_{\beta \rightarrow \alpha} S_\beta^\epsilon(\rho)$ for $\alpha \in \{0, \infty\}$

As in the classical case, this can be written as

$$S_\alpha^\epsilon(\rho) = \begin{cases} \inf_{\sigma \in \mathcal{B}^\epsilon(\rho)} S_\alpha(\sigma), & \text{for } \alpha < 1; \\ \sup_{\sigma \in \mathcal{B}^\epsilon(\rho)} S_\alpha(\sigma), & \text{for } \alpha > 1. \end{cases}$$

Theorem 2.22 *The infimum or supremum in the definition of quantum smooth Rényi entropy is actually a minimum or maximum.*

The proof is the same as in the classical case.

2.6 Ergodicity

Intuitively, an ergodic information source is the most general information source, such that the strong law of large numbers still holds. An independent identically distributed (i.i.d.) source is a special case of an ergodic source.

In this section, the set \mathcal{Z} in the definition of probability distribution may be infinite.

2.6.1 Classical case

The following definition is from [4, page 474-475].

Definition 2.23 (Classical information source) *Define a classical information source by a triple $(\Omega, \mathcal{B}, \mathbb{P})$, with a space Ω (repetitions of an alphabet), algebra of subsets \mathcal{B} and a probability measure \mathbb{P} .*

A random variable X is represented by a function from Ω to \mathbb{C} .

An example for Ω could be $\{0, 1\}^{\mathbb{N}}$ for a source that sends an infinite sequence of bits.

Consider transformations $T : \Omega \rightarrow \Omega$. In the following, these will be time shifts. In the $\{0, 1\}^{\mathbb{N}}$ example, this would remove the first bit from the sequence.

A transformation T is called *stationary* (for a certain information source) if $\mathbb{P}(TA) = \mathbb{P}(A)$ for all $A \in \mathcal{B}$. Intuitively this means the process looks the same at every point in time.

A transformation T is called *ergodic* (for a certain information source) if for every set A with $TA = A$, either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

If T is stationary and ergodic, the process defined by

$$X_n(\omega) = X(T^n\omega)$$

for a random variable X is stationary and ergodic.

We have created some examples of dependent and/or non-identically distributed sources:

- A source that sends alternating zeroes and ones, starting with 0 with 50% probability and 1 with 50% probability.

The probability measure is such that $\mathbb{P}(\omega)$ is $1/2$ for $\omega_1 = [0, 1, 0, 1, \dots]$ and $\omega_2 = [1, 0, 1, 0, \dots]$, and 0 otherwise.

$T(\omega_1) = \omega_2$ and $T(\omega_2) = \omega_1$. T is stationary.

All sets A with $\mathbb{P}(A) > 0$ and $TA = A$ have $\mathbb{P}(A) = 1$, as they must contain ω_1 if they contain ω_2 and vice versa. So T is ergodic.

The random variable X sends each ω to its first element.

This source is ergodic.

- A source with the following probability table per bit:

1/4	send 0 and send only zeroes from this point on
1/4	send 1 and send only ones from this point on
1/4	send 0 and continue following this table
1/4	send 1 and continue following this table

Take $\omega_1 = [0, 1, 1, 1, \dots]$ and $\omega_2 = [1, 1, 1, 1, \dots]$ (both continue with ones forever). Then $T(\omega_1) = \omega_2$. Note that $\mathbb{P}(\omega_2) = 1/4 + 1/4P(\omega_2)$ and $\mathbb{P}(\omega_1) = 1/4P(\omega_2)$. So the transformation T is not stationary for this probability measure.

The transformation T is not ergodic either, $T(\omega_2) = \omega_2$ but $\mathbb{P}(\omega_2) = 1/3$.

This source is not ergodic.

- A source that sends all-ones with probability 1/2 and all-zeroes with probability 1/2.

Take $\omega_1 = [0, 0, 0, 0, \dots]$ and $\omega_2 = [1, 1, 1, 1, \dots]$ (both continue with the same digit forever). So $\mathbb{P}(\omega_1) = \mathbb{P}(\omega_2) = 1/2$.

Then $T(\omega_1) = \omega_1$ and $T(\omega_2) = \omega_2$, so T is stationary for this probability measure, but not ergodic.

This source is not ergodic.

2.6.2 Quantum case

A detailed definition of quantum ergodicity is outside the scope of this thesis; we do not use it directly (instead, we use the AEP). We will just give some background information.

The definition is from [2, section 2].

Some background on discrete quantum information sources (QIS) is needed. The definition contains three main components.

The first component is an algebra, corresponding to Ω and \mathcal{B} in the classical case. As building block we will use the algebra $\mathcal{A} = \mathcal{B}(\mathcal{H})$, the linear operators on the finite dimensional Hilbert space \mathcal{H} . If the QIS emits qubits, \mathcal{H} will be \mathbb{C}^2 . One can also use other choices for \mathcal{A} but we do not need this.

For a finite subset $\Lambda \subset \mathbb{Z}$ the “local” algebra \mathcal{A}_Λ is given by $\mathcal{A}_\Lambda := \bigotimes_{z \in \Lambda} \mathcal{A}_z$. Then the quasilocal algebra \mathcal{A}^∞ is defined as the operator norm closure of the local *-algebra $\mathcal{A}_{\text{loc}} := \bigcup_{\Lambda \subset \mathbb{Z}} \mathcal{A}_\Lambda$.

The second component corresponds to \mathbb{P} in the classical case. A state on the quasilocal algebra is given by a normed positive functional Ψ , i.e. $\Psi(\mathbf{1}) = 1$ and $\Psi(A) \geq 0$ for all $A \in \mathcal{A}^\infty$ with $A \geq 0$.

The third component corresponds to T in the classical case. The shift T is defined on \mathcal{A}_{loc} as follows. For integers $z_1 \leq z_2$ and $\Lambda := \{z_1, z_1 + 1, \dots, z_2\}$ ($\Lambda + 1 = \{z_1 + 1, \dots, z_2 + 1\}$)

$$T : \mathcal{A}_\Lambda \rightarrow \mathcal{A}_{\Lambda+1}, a \simeq a \otimes \mathbf{1} \mapsto T(a) = \mathbf{1} \otimes a \simeq a.$$

The canonical extension of T onto \mathcal{A}^∞ is an $*$ -automorphism on \mathcal{A}^∞ .

Definition 2.24 (Quantum information source) *The triple $(\mathcal{A}^\infty, \Psi, T)$ defines a quantum dynamical system. This is the mathematical model for a discrete QIS.*

$(\mathcal{A}^\infty, \Psi, T)$ is called *stationary* if for all $a \in \mathcal{A}^\infty$ it holds that $\Psi(Ta) = \Psi(a)$.

We will deal only with stationary QIS, hence we can assume without loss of generality that all integer intervals are of the form $\Lambda = \{1, \dots, n\}$ with $n \geq 1$. We write $\rho^{(n)}$ instead of $\rho^{(\Lambda)}$ and $\Psi^{(n)}$ instead of $\Psi^{(\Lambda)}$.

A stationary QIS $(\mathcal{A}^\infty, \Psi, T)$ is *ergodic* if

$$\lim_{n \rightarrow \infty} \Psi \left(\left(\frac{1}{n} \sum_{i=0}^{n-1} T^i(a) \right)^2 \right) = \Psi(a)^2$$

for all self-adjoint $a \in \mathcal{A}^\infty$.

2.7 Asymptotic equipartition property

The asymptotic equipartition property (AEP) says that given enough repetitions of a stationary ergodic distribution, most probability mass is in the “typical set” where every possibility has approximately the same probability. The AEP is also known as the Shannon-McMillan theorem.

In this section, the set \mathcal{Z} in the definition of probability distribution may no longer be infinite.

2.7.1 Classical case

Let \mathbb{P} be a stationary ergodic distribution over $\mathcal{Z} = \{0, 1\}$.

Definition 2.25 (Entropy rate) *Entropy rate is the average per symbol Shannon entropy:*

$$h(\mathbb{P}) := \lim_{n \rightarrow \infty} \frac{1}{n} H(\mathbb{P}^n)$$

Definition 2.26 (Typical sequences, typical set) *A sequence $z^n \in \{0, 1\}^n$ is called ϵ -typical if*

$$2^{-n(h(\mathbb{P})+\epsilon)} \leq \mathbb{P}(z^n) \leq 2^{-n(h(\mathbb{P})-\epsilon)}$$

The typical set T_ϵ^n is the set of all ϵ -typical sequences.

Theorem 2.27 (Classical AEP) *Given $\epsilon > 0$, there is an $N \in \mathbb{N}$ such that for all $n > N$, $\mathbb{P}(T_\epsilon^n) \geq 1 - \epsilon$.*

2.7.2 Quantum case

This theorem is from [2, section 5].

Let $(\mathcal{A}^\infty, \Psi, T)$ be a stationary QIS. Using the one-to-one correspondence between a stationary Ψ and a family of density operators $\{\rho^{(n)} | n \in \mathbb{N}\}$, an average per-symbol entropy can be defined:

Definition 2.28 (Entropy rate of a QIS) $s(\rho) := \lim_{n \rightarrow \infty} \frac{1}{n} S(\rho^{(n)})$

We call certain states “typical”.

Definition 2.29 (Typical state, typical subspace) A pure state $|e_i^{(n)}\rangle$, where $|e_i^{(n)}\rangle$ is an eigenvector of $\rho^{(n)}$, is called ϵ -typical if the corresponding eigenvalue $\lambda_i^{(n)}$ satisfies

$$2^{-n(s(\rho)+\epsilon)} \leq \lambda_i^{(n)} \leq 2^{-n(s(\rho)-\epsilon)}.$$

The typical subspace $\mathcal{T}_\epsilon^{(n)}$ is the linear hull of all ϵ -typical states.

Theorem 2.30 (Quantum AEP) Let $\epsilon > 0$ and let ρ be a stationary ergodic QIS with local densities $\rho^{(n)}$. Then there exists an $N \in \mathbb{N}$ such that for all $n \geq N$

1. $\text{tr}(\rho^{(n)} P_{\mathcal{T}_\epsilon^{(n)}}) \geq 1 - \epsilon$, where $P_{\mathcal{T}_\epsilon^{(n)}}$ is the projector onto the subspace $\mathcal{T}_\epsilon^{(n)}$.
2. $\text{tr}(P_{\mathcal{T}_\epsilon^{(n)}}) \leq 2^{n(s(\rho)+\epsilon)}$

Chapter 3

A closer look at smooth Rényi entropy

3.1 Comparison of the two definitions, classical case

This section compares the two definitions of smooth Rényi entropy (different balls), for the classical case.

In the rest of this section, $H_{o,\alpha}^\epsilon$ stands for smooth Rényi entropy using the statistical distance ball, and $H_{n,\alpha}^\epsilon$ for smooth Rényi entropy using the truncation ball.

For $\alpha = 1$ or $\epsilon = 0$, these are both equal to the Shannon entropy; we will not allow these cases in the following proofs.

3.1.1 Maps between the balls

Several simple maps between $\mathcal{B}_o^\epsilon(\mathbb{P})$ and $\mathcal{B}_n^\epsilon(\mathbb{P})$ can be defined.

Definition 3.1 Define the map $F_{\text{cut}} : \mathcal{B}_o^\epsilon(\mathbb{P}) \rightarrow \mathcal{B}_n^\epsilon(\mathbb{P})$:

Given a \mathbb{Q}_o , there is a corresponding \mathbb{Q}_n :

$$\mathbb{Q}_n(z) = \min\{\mathbb{Q}_o(z), \mathbb{P}(z)\}$$

Lemma 3.2 The map F_{cut} is well defined.

Proof From the definition of \mathbb{Q}_o :

$$\frac{1}{2} \left(\sum_{\{z|\mathbb{Q}_o(z)<\mathbb{P}(z)\}} (\mathbb{P}(z) - \mathbb{Q}_o(z)) + \sum_{\{z|\mathbb{Q}_o(z)>\mathbb{P}(z)\}} (\mathbb{Q}_o(z) - \mathbb{P}(z)) \right) \leq \epsilon$$

so from $\sum_z \mathbb{Q}_o(z) = 1$

$$\sum_{\{z|\mathbb{Q}_o(z)<\mathbb{P}(z)\}} (\mathbb{P}(z) - \mathbb{Q}_o(z)) = \sum_{\{z|\mathbb{Q}_o(z)>\mathbb{P}(z)\}} (\mathbb{Q}_o(z) - \mathbb{P}(z)) \leq \epsilon$$

and

$$\sum_z \mathbb{Q}_n(z) \geq 1 - \epsilon. \quad \square$$

Definition 3.3 Define the map $F_{\text{norm}} : \mathcal{B}_n^\epsilon(\mathbb{P}) \rightarrow \mathcal{B}_0^\epsilon(\mathbb{P})$:

Given a \mathbb{Q}_n , there is a corresponding \mathbb{Q}_0 (\mathbb{Q}_{n^*}) as follows:

$$\mathbb{Q}_{n^*}(z) = \frac{\mathbb{Q}_n(z)}{\sum_x \mathbb{Q}_n(x)}$$

Lemma 3.4 The map F_{norm} is well defined.

Proof Clearly $\sum_z \mathbb{Q}_{n^*}(z) = 1$.

For $\delta(\mathbb{P}, \mathbb{Q}_{n^*})$:

Set $\delta := 1 - \sum_x \mathbb{Q}_n(x)$.

Divide \mathcal{Z} into two parts \mathcal{Z}_1 and \mathcal{Z}_2 such that

$$\begin{aligned} \mathbb{Q}_n(z) < (1 - \delta)\mathbb{P}(z) &\Leftrightarrow z \in \mathcal{Z}_1 \\ \mathbb{Q}_n(z) \geq (1 - \delta)\mathbb{P}(z) &\Leftrightarrow z \in \mathcal{Z}_2 \end{aligned}$$

So $z \in \mathcal{Z}_1 \Leftrightarrow \mathbb{Q}_{n^*}(z) < \mathbb{P}(z)$.

Then

$$\sum_{z \in \mathcal{Z}_1} (\mathbb{P}(z) - \mathbb{Q}_{n^*}(z)) = \sum_{z \in \mathcal{Z}_1} \left(\mathbb{P}(z) - \frac{\mathbb{Q}_n(z)}{1 - \delta} \right) \leq \sum_{z \in \mathcal{Z}_1} (\mathbb{P}(z) - \mathbb{Q}_n(z)) \leq \sum_z (\mathbb{P}(z) - \mathbb{Q}_n(z)) = \delta \leq \epsilon.$$

Since $\sum_z \mathbb{Q}_{n^*}(z) = 1$,

$$\sum_{z \in \mathcal{Z}_1} (\mathbb{P}(z) - \mathbb{Q}_{n^*}(z)) = \sum_{z \in \mathcal{Z}_2} (\mathbb{Q}_{n^*}(z) - \mathbb{P}(z))$$

and

$$\sum_z |\mathbb{P}(z) - \mathbb{Q}_{n^*}(z)|/2 \leq \epsilon. \quad \square$$

Definition 3.5 Define the map $F_{\text{addsmall}} : \mathcal{B}_n^\epsilon(\mathbb{P}) \times \mathbb{N} \rightarrow \mathcal{B}_0^\epsilon(\mathbb{P})$:

Let $\tilde{\mathcal{Z}}$ be a set of M elements of \mathcal{Z} , such that for all $\tilde{z} \in \tilde{\mathcal{Z}}$ it holds that $\mathbb{P}(\tilde{z}) = 0$ (so also $\mathbb{Q}_n(\tilde{z}) = 0$). Given a (\mathbb{Q}_n, M) , there is a corresponding \mathbb{Q}_0 ($\mathbb{Q}_{n^\#}$):

$$\mathbb{Q}_{n^\#}(z) = \begin{cases} \mathbb{Q}_n(z), & z \notin \tilde{\mathcal{Z}} \\ \frac{1 - \sum_x \mathbb{Q}_n(x)}{M}, & z \in \tilde{\mathcal{Z}} \end{cases}$$

3.1.2 Bounds

Using the map F_{cut}

Using theorem 2.20, suppose the minimum/maximum is reached at \mathbb{Q}_o , and $F_{\text{cut}}(\mathbb{Q}_o) = \mathbb{Q}_n$.

$$H_\alpha(\mathbb{Q}_o) = \frac{1}{1-\alpha} \log \sum_z \mathbb{Q}_o(z)^\alpha$$

$$H_\alpha(\mathbb{Q}_n) = \frac{1}{1-\alpha} \log \sum_z \mathbb{Q}_n(z)^\alpha = \frac{1}{1-\alpha} \log \left(\sum_{\mathbb{Q}_o(z) \leq \mathbb{P}(z)} \mathbb{Q}_o(z)^\alpha + \sum_{\mathbb{Q}_o(z) > \mathbb{P}(z)} \mathbb{P}(z)^\alpha \right)$$

If $\alpha = 0$ or $0 < \alpha < 1$:

$$H_\alpha(\mathbb{Q}_n) \leq H_\alpha(\mathbb{Q}_o)$$

$$H_{o,\alpha}^\epsilon(\mathbb{P}) = H_\alpha(\mathbb{Q}_o) \geq H_\alpha(\mathbb{Q}_n) \geq H_{n,\alpha}^\epsilon(\mathbb{P}).$$

If $1 < \alpha < \infty$ or $\alpha = \infty$:

$$H_\alpha(\mathbb{Q}_n) \geq H_\alpha(\mathbb{Q}_o)$$

$$H_{o,\alpha}^\epsilon(\mathbb{P}) = H_\alpha(\mathbb{Q}_o) \leq H_\alpha(\mathbb{Q}_n) \leq H_{n,\alpha}^\epsilon(\mathbb{P}).$$

In both cases, the latter inequality follows from the infimum in the definition of smooth Rényi entropy.

Consider the case $0 < \alpha < 1 \vee 1 < \alpha < \infty$. Note that the inequality between $H_\alpha(\mathbb{Q}_o)$ and $H_\alpha(\mathbb{Q}_n)$ is strict, if $H_{o,\alpha}^\epsilon(\mathbb{P}) \neq H_\alpha(\mathbb{P})$. Furthermore note that clearly $H_{n,\alpha}^\epsilon(\mathbb{P}) \neq H_\alpha(\mathbb{P})$ (because $\epsilon > 0$). So it must be that the inequality between $H_{o,\alpha}^\epsilon(\mathbb{P})$ and $H_{n,\alpha}^\epsilon(\mathbb{P})$ is strict.

A more concrete example of this difference is in section 3.1.3.

Using the map F_{norm}

Using theorem 2.20, suppose the minimum/maximum is reached at \mathbb{Q}_n , and $F_{\text{norm}}(\mathbb{Q}_n) = \mathbb{Q}_{n^*}$.

Set $\delta := 1 - \sum_x \mathbb{Q}_n(x)$. Then $\delta \leq \epsilon$.

Rényi entropy on single points \mathbb{Q}_n and \mathbb{Q}_{n^*} , for $0 < \alpha < \infty$:

$$H_\alpha(\mathbb{Q}_{n^*}) = \frac{1}{1-\alpha} \log \sum_z \mathbb{Q}_{n^*}(z)^\alpha = \frac{1}{1-\alpha} \log \sum_z \mathbb{Q}_n(z)^\alpha - \frac{\alpha}{1-\alpha} \log(1-\delta)$$

For $\alpha = 0$:

$$H_0(\mathbb{Q}_{n^*}) = \log \#\{z | \mathbb{Q}_{n^*}(z) > 0\} = \log \#\{z | \mathbb{Q}_n(z) > 0\} = H_0(\mathbb{Q}_n)$$

For $\alpha = \infty$:

$$\begin{aligned} H_\infty(\mathbb{Q}_{n^*}) &= -\log \max\{z | \mathbb{Q}_{n^*}(z)\} = -\log \max\{z | \frac{1}{1-\delta} \mathbb{Q}_n(z)\} \\ &= -\log \max\{z | \mathbb{Q}_n(z)\} - \log \frac{1}{1-\delta} = -\log \max\{z | \mathbb{Q}_n(z)\} + \log(1-\delta) \end{aligned}$$

If $\alpha = 0$ or $0 < \alpha < 1$:

$$H_{n,\alpha}^\epsilon(\mathbb{P}) - \frac{\alpha}{1-\alpha} \log(1-\epsilon) \geq H_{n,\alpha}^\epsilon(\mathbb{P}) - \frac{\alpha}{1-\alpha} \log(1-\delta) = H_\alpha(\mathbb{Q}_{n^*}) \geq H_{o,\alpha}^\epsilon(\mathbb{P})$$

If $\infty > \alpha > 1$:

$$H_{n,\alpha}^\epsilon(\mathbb{P}) - \frac{\alpha}{1-\alpha} \log(1-\epsilon) \leq H_{n,\alpha}^\epsilon(\mathbb{P}) - \frac{\alpha}{1-\alpha} \log(1-\delta) = H_\alpha(\mathbb{Q}_{n^*}) \leq H_{o,\alpha}^\epsilon(\mathbb{P})$$

If $\alpha = \infty$:

$$H_{n,\alpha}^\epsilon(\mathbb{P}) + \log(1-\epsilon) \leq H_{n,\alpha}^\epsilon(\mathbb{P}) + \log(1-\delta) = H_\alpha(\mathbb{Q}_{n^*}) \leq H_{o,\alpha}^\epsilon(\mathbb{P})$$

In all three cases, the latter inequality follows from the infimum in the definition of smooth Rényi entropy.

Using the map F_{addsmall}

Using theorem 2.20, suppose the minimum/maximum is reached at \mathbb{Q}_n , let $M \in \mathbb{N}$, and $F_{\text{addsmall}}(M, \mathbb{Q}_n) = \mathbb{Q}_{n\#}(M)$.

Set $\delta := 1 - \sum_x \mathbb{Q}_n(x)$. Then $\delta \leq \epsilon$.

Rényi entropy on single points \mathbb{Q}_n and $\mathbb{Q}_{n\#}(M)$, for $1 < \alpha < \infty$:

$$H_\alpha(\mathbb{Q}_{n\#}(M)) = \frac{1}{1-\alpha} \log \sum_z \mathbb{Q}_{n\#}(M)(z)^\alpha = \frac{1}{1-\alpha} \log \left(\sum_z \mathbb{Q}_n(z)^\alpha + M \left(\frac{\delta}{M} \right)^\alpha \right)$$

Note that $\lim_{M \rightarrow \infty} M \left(\frac{\delta}{M} \right)^\alpha = 0$.

Then

$$H_{n,\alpha}^\epsilon(\mathbb{P}) = H_\alpha(\mathbb{Q}_n) = \lim_{M \rightarrow \infty} H_\alpha(\mathbb{Q}_{n\#}(M)) \leq H_{o,\alpha}^\epsilon(\mathbb{P})$$

The latter inequality follows from the infimum in the definition of smooth Rényi entropy.

For $\alpha = \infty$:

Let M be the smallest integer greater than $\frac{\epsilon}{2 - H_\infty(\mathbb{Q}_n)}$. Then $\frac{1 - \sum_x \mathbb{Q}_n(x)}{M} < \max\{z | \mathbb{Q}_n(z)\}$. So

$$H_\infty(\mathbb{Q}_{n\#}(M)) = -\log \max\{z | \mathbb{Q}_{n\#}(M)(z)\} = -\log \max\{z | \mathbb{Q}_n(z)\}.$$

Then it follows that

$$H_{n,\infty}^\epsilon(\mathbb{P}) = H_\infty(\mathbb{Q}_n) = H_\infty(\mathbb{Q}_{n\#}(M)) \leq H_{o,\infty}^\epsilon(\mathbb{P})$$

The latter inequality follows from the infimum in the definition of smooth Rényi entropy.

Together

Summarizing, we get the following bounds. Note that different from [7, full version, section 3.3], we find that there is a difference between the definitions.

If $\alpha = 0$:

$$H_{n,\alpha}^\epsilon(\mathbb{P}) = H_{o,\alpha}^\epsilon(\mathbb{P})$$

If $0 < \alpha < 1$:

$$H_{n,\alpha}^\epsilon(\mathbb{P}) < H_{o,\alpha}^\epsilon(\mathbb{P}) \leq H_{n,\alpha}^\epsilon(\mathbb{P}) - \frac{\alpha}{1-\alpha} \log(1-\epsilon)$$

If $1 < \alpha < \infty$:

$$H_{n,\alpha}^\epsilon(\mathbb{P}) > H_{o,\alpha}^\epsilon(\mathbb{P}) \geq H_{n,\alpha}^\epsilon(\mathbb{P}) - \frac{\alpha}{1-\alpha} \log(1-\epsilon)$$

If $\alpha = \infty$:

$$H_{n,\alpha}^\epsilon(\mathbb{P}) > H_{o,\alpha}^\epsilon(\mathbb{P}) \geq H_{n,\alpha}^\epsilon(\mathbb{P}) + \log(1-\epsilon)$$

The bound for $\alpha = \infty$ and $1 < \alpha < \infty$ can be improved if elements with probability 0 are added to \mathcal{Z} ; for $\alpha = \infty$,

$$H_{n,\alpha}^\epsilon(\mathbb{P}) = H_{o,\alpha}^\epsilon(\mathbb{P})$$

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

$$H_{n,\alpha}^\epsilon(\mathbb{P}) > H_{o,\alpha}^\epsilon(\mathbb{P}) \geq H_{n,\alpha}^\epsilon(\mathbb{P}) - d$$

for any $d > 0$ (the closer d is to 0, the more elements in \mathcal{Z} with probability 0 are needed).

3.1.3 Example of the difference

As an example for the difference between the two definitions, consider $\mathcal{Z} = \{0, 1\}$ and $\mathbb{P}(0) = \mathbb{P}(1) = 1/2$.

For $\alpha = 0$ and $\epsilon < 1/2$, it is clear that for all \mathbb{Q} in either $\mathcal{B}_o^\epsilon(\mathbb{P})$ or $\mathcal{B}_n^\epsilon(\mathbb{P})$, $\mathbb{Q}(0)$ and $\mathbb{Q}(1)$ will both be non-zero. Hence the smooth Rényi entropy is $\log 2$ in both cases.

For $\alpha = 1/2$ and $\epsilon < 1/2$, the best \mathbb{Q} for the truncation ball is $\mathbb{Q}(0) = 1/2 - \epsilon$, $\mathbb{Q}(1) = 1/2$, giving a smooth Rényi entropy of $2 \log(\sqrt{1/2 - \epsilon} + \sqrt{1/2})$. For the statistical distance ball, the best \mathbb{Q} is $\mathbb{Q}(0) = 1/2 - \epsilon$, $\mathbb{Q}(1) = 1/2 + \epsilon$, giving a smooth Rényi entropy of $2 \log(\sqrt{1/2 - \epsilon} + \sqrt{1/2 + \epsilon})$.

For $\alpha = \infty$ and $\epsilon < 1/4$, the best result in the truncation ball is lowering both $\mathbb{Q}(0)$ and $\mathbb{Q}(1)$ to $1/2 - \epsilon/2$, giving a smooth Rényi entropy of $-\log(1/2 - \epsilon/2)$. However, in the statistical distance ball no improvement is possible, leaving the smooth Rényi entropy at $\log 2$.

If we instead use $\mathcal{Z} = \{0, 1, 2\}$ and take $\mathbb{P}(2) = 0$, the difference between the statistical distance ball and the truncation ball decreases. For $\alpha = \infty$ and ϵ small enough, the extra probability mass can be put in $\mathbb{Q}(2)$, giving $\mathbb{Q}(0) = \mathbb{Q}(1) = 1/2 - \epsilon/2$ and $\mathbb{Q}(2) = \epsilon$ and a smooth Rényi entropy of $-\log(1/2 - \epsilon/2)$.

3.2 Comparison of the two definitions, quantum case

This section will generalise the previous section to quantum information.

3.2.1 Comparison between quantum and classical information

We need some lemmas about the relation between classical probability distributions and density matrices.

Lemma 3.6 (Weyl's Monotonicity Theorem) *If A, B are n by n Hermitian, and B is positive, then $\lambda_i(A) \leq \lambda_i(A + B)$ for all $i = 1, \dots, n$, where $\lambda_i(M)$ is the i 'th eigenvalue (ordered from largest to smallest) of the Hermitian matrix M .*

Proof See for example [1, Corollary III.2.3].

Lemma 3.7 *Let ρ be a density operator with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.*

1. *Given a matrix σ with eigenvalues $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$,*

$$\sigma \in \mathcal{B}_n^\epsilon(\rho) \Rightarrow \underline{\mu} \in \mathcal{B}_n^\epsilon(\underline{\lambda})$$

2. *Given real numbers μ_1, \dots, μ_n such that $\underline{\mu} \in \mathcal{B}_n^\epsilon(\underline{\lambda})$, there exists a matrix σ with eigenvalues μ_1, \dots, μ_n such that $\sigma \in \mathcal{B}_n^\epsilon(\rho)$.*

Proof Recall that

$$\mathcal{B}_n^\epsilon(\underline{\lambda}) = \{\underline{\mu} \mid \forall_i \mu_i \leq \lambda_i, \sum_i \mu_i \geq 1 - \epsilon, \forall_i \mu_i \geq 0\}$$

and

$$\mathcal{B}_n^\epsilon(\rho) = \{\sigma \mid \sigma \leq \rho, \text{tr}(\sigma) \geq 1 - \epsilon, \sigma \geq 0\}$$

1. Let σ be a matrix with eigenvalues $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$ and suppose $\sigma \in \mathcal{B}_n^\epsilon(\rho)$.

σ positive gives $\forall_i \mu_i \geq 0$. $\rho - \sigma$ is positive so $\lambda_i \geq \mu_i$ for all i (from lemma 3.6). $\text{tr}(\sigma) \geq 1 - \epsilon$ gives $\sum_i \mu_i \geq 1 - \epsilon$.

So $\underline{\mu}$ is in the classical truncation ball around $\underline{\lambda}$.

2. Because ρ is Hermitian, it is diagonalisable: there are v_i ($i = 1, \dots, n$) such that

$$\rho = \sum_i \lambda_i |v_i\rangle\langle v_i|.$$

Let σ be

$$\sigma := \sum_i \mu_i |v_i\rangle\langle v_i|.$$

$\forall_i \mu_i \geq 0$ so σ is positive.

The eigenvalues of $\rho - \sigma = \sum_i (\lambda_i - \mu_i) |v_i\rangle\langle v_i|$ are $\lambda_i - \mu_i$, which are non-negative real, so $\rho - \sigma$ is positive.

$\sum_i \mu_i \geq 1 - \epsilon$ so $\text{tr}(\sigma) \geq 1 - \epsilon$.

So σ is in the quantum truncation ball around ρ .

□

Lemma 3.8 *If $\lambda_1, \dots, \lambda_n$ are the eigenvalues of the density matrix ρ ,*

$$S_{n,\alpha}^\epsilon(\rho) = H_{n,\alpha}^\epsilon(\underline{\lambda}).$$

Proof

Recall from the definitions that $S_\alpha(\rho) = H_\alpha(\underline{\lambda})$.

$$S_\alpha^\epsilon(\rho) = \inf_{\sigma \in \mathcal{B}^\epsilon(\rho)} S_\alpha(\sigma) = \inf_{\underline{\mu} \in \mathcal{B}^\epsilon(\underline{\lambda})} H_\alpha(\underline{\mu}) = H_\alpha^\epsilon(\underline{\lambda})$$

□

Lemma 3.9 *Given two density operators ρ and σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ and $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$,*

$$\delta(\rho, \sigma) \geq \frac{1}{2} \sum_i |\lambda_i - \mu_i|.$$

Proof

There is a unitary matrix U and a diagonal matrix D such that $\rho - \sigma = UDU^\dagger$.

There are positive diagonal matrices D^+ and D^- such that $D = D^+ - D^-$.

Then $\rho - \sigma = UDU^\dagger = U(D^+ - D^-)U^\dagger = UD^+U^\dagger - UD^-U^\dagger =: Q - S$. Q and S are positive matrices with support on orthogonal spaces. So $|\rho - \sigma| = Q + S$.

Define the matrix $V := S + \rho = Q + \sigma$. This is a positive matrix.

Let $\nu_1 \geq \nu_2 \geq \dots \geq \nu_n$ be the eigenvalues of V . It follows from lemma 3.6 that $\nu_i \geq \max\{\lambda_i, \mu_i\}$ and $\nu_i \geq \frac{1}{2}\lambda_i + \frac{1}{2}\mu_i + \frac{1}{2}|\lambda_i - \mu_i|$.

Then $\delta(\rho, \sigma) = \frac{1}{2}(\text{tr}(Q) + \text{tr}(S))/2 = \text{tr}(V) - \text{tr}(\rho) - \text{tr}(\sigma) = \sum_i (\nu_i - \frac{1}{2}\lambda_i - \frac{1}{2}\mu_i) \geq \frac{1}{2} \sum_i |\lambda_i - \mu_i|$.

□

Lemma 3.10 *Let ρ be a density operator with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.*

1. *Given a matrix σ with eigenvalues $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$,*

$$\sigma \in \mathcal{B}_0^\epsilon(\rho) \Rightarrow \underline{\mu} \in \mathcal{B}_0^\epsilon(\underline{\lambda})$$

2. *Given real numbers μ_1, \dots, μ_n such that $\underline{\mu} \in \mathcal{B}_0^\epsilon(\underline{\lambda})$, there exists a matrix σ with eigenvalues μ_1, \dots, μ_n such that $\sigma \in \mathcal{B}_0^\epsilon(\rho)$.*

Proof

Recall that

$$\mathcal{B}_0^\epsilon(\underline{\lambda}) = \{\underline{\mu} \mid \delta(\underline{\lambda}, \underline{\mu}) \leq \epsilon, \sum_i \mu_i = 1, \forall_i \mu_i \geq 0\}$$

and

$$\mathcal{B}_0^\epsilon(\rho) = \{\sigma \mid \delta(\rho, \sigma) \leq \epsilon, \text{tr}(\sigma) = 1, \sigma \geq 0\}$$

1. Let $\sigma \in \mathcal{B}_0^\epsilon(\rho)$. Then $\delta(\rho, \sigma) \leq \epsilon$ so $\frac{1}{2} \sum_i |\lambda_i - \mu_i| \leq \epsilon$ (from lemma 3.9). Also σ is a density operator so $\forall_i \mu_i \geq 0$ and $\sum_i \mu_i = 1$. So $\underline{\mu} \in \mathcal{B}_0^\epsilon(\underline{\lambda})$.
2. Because ρ is Hermitian, it is diagonalisable: there are v_i ($i = 1, \dots, n$) such that

$$\rho = \sum_i \lambda_i |v_i\rangle\langle v_i|.$$

Let σ be

$$\sigma := \sum_i \mu_i |v_i\rangle\langle v_i|.$$

$\forall_i \mu_i \geq 0$ so σ is positive.

$\sum_i \mu_i = 1$ so $\text{tr}(\sigma) = 1$.

Because the ρ and σ diagonalise in the same way, the trace distance between them is equal to the statistical distance between the eigenvalues, so $\delta(\rho, \sigma) = \delta(\underline{\lambda}, \underline{\mu}) \leq \epsilon$.

So σ is in the quantum “old” ball around ρ .

□

Lemma 3.11 *If $\lambda_1, \dots, \lambda_n$ are the eigenvalues of the density matrix ρ ,*

$$S_{o,\alpha}^\epsilon(\rho) = H_{o,\alpha}^\epsilon(\underline{\lambda}).$$

Proof

This follows immediately from the previous lemma and the alternate characterization of quantum smooth Rényi entropy ($S_\alpha(\rho) = H_\alpha(\underline{\lambda})$). \square

Then we can construct analogons for F_{cut} and F_{norm} , in such a way that the eigenvalues are transformed like the probabilities in the classical case. These are not necessary for the proof of equivalency.

Definition 3.12 *Define the map $F_{\text{cut}} : \mathcal{B}_o^\epsilon(\rho) \rightarrow \mathcal{B}_n^\epsilon(\rho)$,*

$$\sigma_o \mapsto \sigma_n := \rho - \text{Pos}(\rho - \sigma)$$

with

$$\text{Pos}(\rho - \sigma) = \sum_i \max\{\nu_i, 0\} |v_i\rangle\langle v_i|$$

if

$$\rho - \sigma = \sum_i \nu_i |v_i\rangle\langle v_i|.$$

Note that $\rho - \sigma$ is diagonalisable because it is the difference of two Hermitian matrices.

Definition 3.13 *Define the map $F_{\text{norm}} : \mathcal{B}_n^\epsilon(\rho) \rightarrow \mathcal{B}_o^\epsilon(\rho)$,*

$$\sigma_n \mapsto \sigma_{n^*} := \frac{\sigma_n}{\text{tr}(\sigma_n)}.$$

3.2.2 Bounds

Apply lemmas 3.8 and 3.11 to the result in the classical case, this gives:

If $\alpha = 0$ or $0 < \alpha < 1$:

$$S_{n,\alpha}^\epsilon(\rho) - \frac{\alpha}{1-\alpha} \log(1-\epsilon) \geq S_{o,\alpha}^\epsilon(\rho) \geq S_{n,\alpha}^\epsilon(\rho)$$

If $\infty > \alpha > 1$:

$$S_{n,\alpha}^\epsilon(\rho) - \frac{\alpha}{1-\alpha} \log(1-\epsilon) \leq S_{o,\alpha}^\epsilon(\rho) \leq S_{n,\alpha}^\epsilon(\rho)$$

If $\alpha = \infty$:

$$S_{n,\alpha}^\epsilon(\rho) + \log(1-\epsilon) \leq S_{o,\alpha}^\epsilon(\rho) \leq S_{n,\alpha}^\epsilon(\rho)$$

Similarly to the classical case, the latter two bounds can be improved if eigenvectors with eigenvalue 0 are added to ρ . For $\infty > \alpha > 1$ this brings the entropies arbitrarily close; for $\alpha = \infty$ it makes them equal with one such eigenvalue if ϵ is sufficiently small.

3.3 Relations between smooth Rényi entropy of different orders

In this section we will derive inequalities between smooth Rényi entropy of different orders. In the next chapter we will do the proof for orders 0 and ∞ only (for order 1, Smooth Rényi entropy is equal to Shannon/Von Neumann entropy).

From [6, section 2.4], for $\alpha, \beta \in [0, \infty]$ and ρ a density matrix

$$\alpha \leq \beta \Leftrightarrow S_\alpha(\rho) \geq S_\beta(\rho)$$

Now let σ be a truncated density matrix (trace not necessarily equal to 1). Then

$$\begin{aligned} S_\alpha(\sigma) &= S_\alpha\left(\frac{\sigma}{\text{tr}(\sigma)}\text{tr}(\sigma)\right) = \frac{1}{1-\alpha} \log\left(\text{tr}\left(\left(\frac{\sigma}{\text{tr}(\sigma)}\right)^\alpha\right) (\text{tr}(\sigma))^\alpha\right) \\ &= S_\alpha\left(\frac{\sigma}{\text{tr}(\sigma)}\right) + \frac{1}{1-\alpha} \log(\text{tr}(\sigma)). \end{aligned}$$

Combining these two formulas and reordering the terms gives

$$S_\alpha(\sigma) - \frac{\alpha}{1-\alpha} \log(\text{tr}(\sigma)) = S_\alpha\left(\frac{\sigma}{\text{tr}(\sigma)}\right) \geq S_\beta\left(\frac{\sigma}{\text{tr}(\sigma)}\right) = S_\beta(\sigma) - \frac{\beta}{1-\beta} \log(\text{tr}(\sigma))$$

which can be simplified to

$$S_\alpha(\sigma) + \frac{1}{1-\alpha} \log(\text{tr}(\sigma)) = S_\alpha\left(\frac{\sigma}{\text{tr}(\sigma)}\right) \geq S_\beta\left(\frac{\sigma}{\text{tr}(\sigma)}\right) = S_\beta(\sigma) + \frac{1}{1-\beta} \log(\text{tr}(\sigma)).$$

Note that $1 - \epsilon \leq \text{tr}(\sigma) \leq 1$.

From this and theorem 2.22, results for smooth Rényi entropy with the truncation ball can be derived:

For $\alpha < \beta < 1$:

$$\begin{aligned} & S_\alpha^\epsilon(\rho) + \frac{1}{1-\alpha} \log(\text{tr}(\sigma)) \\ &= S_\alpha(\sigma) + \frac{1}{1-\alpha} \log(\text{tr}(\sigma)) \\ &\geq S_\beta(\sigma) + \frac{1}{1-\beta} \log(\text{tr}(\sigma)) \\ &\geq S_\beta^\epsilon(\rho) + \frac{1}{1-\beta} \log(\text{tr}(\sigma)) \end{aligned}$$

For $1 < \alpha < \beta$:

$$\begin{aligned} & S_\beta^\epsilon(\rho) + \frac{1}{1-\beta} \log(\text{tr}(\sigma)) \\ &= S_\beta(\sigma) + \frac{1}{1-\beta} \log(\text{tr}(\sigma)) \\ &\leq S_\alpha(\sigma) + \frac{1}{1-\alpha} \log(\text{tr}(\sigma)) \\ &\leq S_\alpha^\epsilon(\rho) + \frac{1}{1-\alpha} \log(\text{tr}(\sigma)) \end{aligned}$$

Furthermore, we have the following theorem:

Theorem 3.14 1. If $\alpha < 1$, then

$$S_0^{2\epsilon}(\rho) \leq S_\alpha^\epsilon(\rho) + \frac{\log(1/\epsilon)}{1-\alpha}$$

2. If $\alpha > 1$, then

$$S_{\infty}^{2\epsilon}(\rho) \geq S_{\alpha}^{\epsilon}(\rho) - \frac{\log(1/\epsilon)}{\alpha - 1}$$

Proof Apply lemma 3.11 to [8, lemma 2] and take $\epsilon = \epsilon'$. □

Together, it follows that proving the limit result for $\alpha = 0$, $\alpha = 1$ and $\alpha = \infty$ implies the limit result for general α .

Chapter 4

Asymptotic results for smooth Rényi entropy

In this section we will prove that Smooth Rényi entropy goes to Shannon/Von Neumann entropy in the limit for $n \rightarrow \infty$ and $\epsilon \rightarrow 0$.

4.1 Definitions (classical case)

Let \mathbb{P} be a stationary ergodic probability distribution on $\mathcal{Z} = \{0, 1\}^{\mathbb{N}}$. Define

$$h(\mathbb{P}) = \lim_{n \rightarrow \infty} -\frac{1}{n} \sum_{z^n} \mathbb{P}(z^n) \log(\mathbb{P}(z^n))$$

to be the limit of the Shannon entropy for $n \rightarrow \infty$.

Define $h_\alpha(\mathbb{P})$ to be the limit of the smooth Rényi entropy for $\epsilon \rightarrow 0$ and $n \rightarrow \infty$:

$$\begin{aligned} h_0^\epsilon(\mathbb{P}) &= \lim_{n \rightarrow \infty} \frac{1}{n} H_0^{\epsilon, n}(\mathbb{P}) \\ h_\infty^\epsilon(\mathbb{P}) &= \lim_{n \rightarrow \infty} \frac{1}{n} H_\infty^{\epsilon, n}(\mathbb{P}) \end{aligned}$$

4.2 Classical case, truncation ball

$\mathcal{B}^\epsilon(\mathbb{P})$ is the truncation ball.

Theorem 4.1 $h_0^\epsilon(\mathbb{P})$ is close to $h(\mathbb{P})$: for all $0 < \epsilon < \frac{1}{2}$:

1. $h_0^\epsilon(\mathbb{P}) \leq h(\mathbb{P}) + \epsilon$
2. $h_0^\epsilon(\mathbb{P}) \geq h(\mathbb{P}) - 2\epsilon$

Proof Apply theorem 2.27 and let T_ϵ^n be the typical set. For all $z^n \in T_\epsilon^n$, $2^{-n(h(\mathbb{P})+\epsilon)} \leq \mathbb{P}(z^n) \leq 2^{-n(h(\mathbb{P})-\epsilon)}$, and $2^{n(h(\mathbb{P})+\epsilon)} \leq T_\epsilon^n \leq 2^{n(h(\mathbb{P})-\epsilon)}$.

1. Define the function

$$\mathbb{P}^{\epsilon,n}(z^n) = \begin{cases} \mathbb{P}(z^n) & \text{if } z^n \in T_\epsilon^n, \\ 0 & \text{if } z^n \notin T_\epsilon^n. \end{cases}$$

From the AEP it follows that $\mathbb{P}^{\epsilon,n}(T_\epsilon^n) \geq 1 - \epsilon$ for n sufficiently large. Also, clearly $\mathbb{P}^{\epsilon,n}(z^n) \leq \mathbb{P}(z^n)$ for all z^n . So $\mathbb{P}^{\epsilon,n}(z^n) \in \mathcal{B}^{\epsilon,n}(\mathbb{P})$.

$$H_0^{\epsilon,n}(\mathbb{P}^{\epsilon,n}) \leq \log 2^{n(h(\mathbb{P})+\epsilon)} = n(h(\mathbb{P}) + \epsilon)$$

So

$$h_0^\epsilon(\mathbb{P}) = \lim_{n \rightarrow \infty} \frac{1}{n} \inf_{\mathbb{Q}} H_0^n(\mathbb{Q}) \leq \lim_{n \rightarrow \infty} \frac{1}{n} n(h(\mathbb{P}) + \epsilon) = \lim_{n \rightarrow \infty} (h(\mathbb{P}) + \epsilon) \leq h(\mathbb{P}) + \epsilon.$$

2. Let $\mathbb{Q} \in \mathcal{B}^{\epsilon,n}(\mathbb{P})$.

From the definition: $\sum_{z^n} \mathbb{Q}(z^n) \geq 1 - \epsilon$

$$\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n) + \sum_{z^n \notin T_\epsilon^n} \mathbb{Q}(z^n) \geq 1 - \epsilon$$

$$\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n) \geq 1 - 2\epsilon.$$

If n is large enough and $\epsilon < 1/2$, $1 - 2\epsilon \geq 2^{-n\epsilon}$.

Then

$$\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n) \geq 1 - 2\epsilon \geq 2^{-n\epsilon} = 2^{n(h(\mathbb{P})-2\epsilon)} 2^{-n(h(\mathbb{P})-\epsilon)} \geq 2^{n(h(\mathbb{P})-2\epsilon)} \max_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n)$$

so

$$2^{n(h(\mathbb{P})-2\epsilon)} \leq \frac{\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n)}{\max_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n)} = \frac{\sum_{z^n \in T_\epsilon^n, \mathbb{Q}(z^n) > 0} \mathbb{Q}(z^n)}{\max_{z^n \in T_\epsilon^n, \mathbb{Q}(z^n) > 0} \mathbb{Q}(z^n)} \leq \#\{z^n \in T_\epsilon^n | \mathbb{Q}(z^n) > 0\}$$

and

$$\log \#\{z^n \in T_\epsilon^n | \mathbb{Q}(z^n) > 0\} \geq n(h(\mathbb{P}) - 2\epsilon).$$

So for all $\mathbb{Q} \in \mathcal{B}^{\epsilon,n}(\mathbb{P})$, we have $H_0^n(\mathbb{Q}) > n(h(\mathbb{P}) - 2\epsilon)$. Hence $H_0^{\epsilon,n}(\mathbb{P}) > n(h(\mathbb{P}) - 2\epsilon)$ and $h_0^\epsilon(\mathbb{P}) \geq h(\mathbb{P}) - 2\epsilon$.

□

Theorem 4.2 $h_\infty^\epsilon(\mathbb{P})$ is close to $h(\mathbb{P})$: for all $0 < \epsilon < \frac{1}{2}$:

1. $h_\infty^\epsilon(\mathbb{P}) \geq h(\mathbb{P}) - \epsilon$
2. $h_\infty^\epsilon(\mathbb{P}) \leq h(\mathbb{P}) + 2\epsilon$

Proof Apply theorem 2.27 and let T_ϵ^n be the typical set. For all $z^n \in T_\epsilon^n$, $2^{-n(h(\mathbb{P})+\epsilon)} \leq \mathbb{P}(z^n) \leq 2^{-n(h(\mathbb{P})-\epsilon)}$, and $2^{n(h(\mathbb{P})+\epsilon)} \leq T_\epsilon^n \leq 2^{n(h(\mathbb{P})-\epsilon)}$.

1. Define the function

$$\mathbb{P}^{\epsilon,n}(z^n) = \begin{cases} \mathbb{P}(z^n) & \text{if } z^n \in T_\epsilon^n, \\ 0 & \text{if } z^n \notin T_\epsilon^n. \end{cases}$$

From the AEP it follows that $\mathbb{P}^{\epsilon,n}(T_\epsilon^n) \geq 1 - \epsilon$ for n sufficiently large. Also, clearly $\mathbb{P}^{\epsilon,n}(z^n) \leq \mathbb{P}(z^n)$ for all z^n . So $\mathbb{P}^{\epsilon,n}(z^n) \in \mathcal{B}^{\epsilon,n}(\mathbb{P})$.

$$H_\infty^{\epsilon,n}(\mathbb{P}^{\epsilon,n}) \geq -\log 2^{-n(h(\mathbb{P})-\epsilon)} = n(h(\mathbb{P}) - \epsilon)$$

So

$$h_\infty^\epsilon(\mathbb{P}) = \lim_{n \rightarrow \infty} \frac{1}{n} \inf_{\mathbb{Q}} H_\infty^n(\mathbb{Q}) \geq \lim_{n \rightarrow \infty} \frac{1}{n} n(h(\mathbb{P}) - \epsilon) = \lim_{n \rightarrow \infty} (h(\mathbb{P}) - \epsilon) \geq h(\mathbb{P}) - \epsilon.$$

2. Let $\mathbb{Q} \in \mathcal{B}^{\epsilon,n}(\mathbb{P})$.

From the definition: $\sum_{z^n} \mathbb{Q}(z^n) \geq 1 - \epsilon$

$$\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n) + \sum_{z^n \notin T_\epsilon^n} \mathbb{Q}(z^n) \geq 1 - \epsilon$$

$$\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n) \geq 1 - 2\epsilon.$$

If n is large enough and $\epsilon < 1/2$, $1 - 2\epsilon \geq 2^{-n\epsilon}$.

Then

$$\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n) \geq 1 - 2\epsilon \geq 2^{-n\epsilon} = 2^{-n(h(\mathbb{P})+2\epsilon)} 2^{n(h(\mathbb{P})+\epsilon)} \geq 2^{-n(h(\mathbb{P})+2\epsilon)} \#T_\epsilon^n$$

so

$$2^{-n(h(\mathbb{P})+2\epsilon)} \leq \frac{\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n)}{\#T_\epsilon^n} \leq \max_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n) \leq \max_{z^n} \mathbb{Q}(z^n)$$

and

$$-\log \max_{z^n} \mathbb{Q}(z^n) \leq n(h(\mathbb{P}) + 2\epsilon).$$

So for all $\mathbb{Q} \in \mathcal{B}^{\epsilon,n}(\mathbb{P})$, we have $H_\infty^n(\mathbb{Q}) \leq n(h(\mathbb{P}) + 2\epsilon)$. Hence $H_\infty^{\epsilon,n}(\mathbb{P}) \leq n(h(\mathbb{P}) + 2\epsilon)$ and $h_\infty^\epsilon(\mathbb{P}) \leq h(\mathbb{P}) + 2\epsilon$.

□

4.3 Classical case, statistical distance ball

Theorem 4.3 $\tilde{h}_0^\epsilon(\mathbb{P})$ is close to $h(\mathbb{P})$: for all $0 < \epsilon < \frac{1}{2}$:

1. $\tilde{h}_0^\epsilon(\mathbb{P}) \leq h(\mathbb{P}) + \epsilon$
2. $\tilde{h}_0^\epsilon(\mathbb{P}) \geq h(\mathbb{P}) - 2\epsilon$

Proof Apply theorem 2.27 and let T_ϵ^n be the typical set. For all $z^n \in T_\epsilon^n$, $2^{-n(h(\mathbb{P})+\epsilon)} \leq \mathbb{P}(z^n) \leq 2^{-n(h(\mathbb{P})-\epsilon)}$, and $2^{n(h(\mathbb{P})+\epsilon)} \leq T_\epsilon^n \leq 2^{n(h(\mathbb{P})-\epsilon)}$.

1. Define the distribution

$$\mathbb{P}^{\epsilon,n}(z^n) = \begin{cases} \mathbb{P}(z^n)/\mathbb{P}(T_\epsilon^n) & \text{if } z^n \in T_\epsilon^n, \\ 0 & \text{if } z^n \notin T_\epsilon^n. \end{cases}$$

From comparison of definitions, $F_{\text{norm}}, \mathbb{P}^{\epsilon,n}(z^n) \in \mathcal{B}^{\epsilon,n}(\mathbb{P})$.

$$\tilde{H}_0^{\epsilon,n}(\mathbb{P}^{\epsilon,n}) \leq \log 2^{n(h(\mathbb{P})+\epsilon)} = n(h(\mathbb{P}) + \epsilon)$$

So

$$\tilde{h}_0^\epsilon(\mathbb{P}) = \lim_{n \rightarrow \infty} \frac{1}{n} \inf_{\mathbb{Q}} \tilde{H}_0^n(\mathbb{Q}) \leq \lim_{n \rightarrow \infty} \frac{1}{n} n(h(\mathbb{P}) + \epsilon) = \lim_{n \rightarrow \infty} (h(\mathbb{P}) + \epsilon) \leq h(\mathbb{P}) + \epsilon.$$

2. Let $\mathbb{Q} \in \mathcal{B}^{\epsilon,n}(\mathbb{P})$.

From the definition: $\sum_{z^n} \mathbb{Q}(z^n) \geq 1 - \epsilon$

$$\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n) + \sum_{z^n \notin T_\epsilon^n} \mathbb{Q}(z^n) \geq 1 - \epsilon$$

$$\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n) \geq 1 - 2\epsilon.$$

If n is large enough and $\epsilon < 1/2$, $1 - 2\epsilon \geq 2^{-n\epsilon}$.

Now change \mathbb{Q} : for every z^n with $\mathbb{Q}(z^n) > \mathbb{P}(z^n)$, replace $\mathbb{Q}(z^n)$ by $\mathbb{P}(z^n)$. This changed \mathbb{Q} is in the truncation ball (again see comparison of definitions, F_{cut}).

Then the same proof applies:

$$\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n) \geq 1 - 2\epsilon \geq 2^{-n\epsilon} = 2^{n(h(\mathbb{P})-2\epsilon)} 2^{-n(h(\mathbb{P})-\epsilon)} \geq 2^{n(h(\mathbb{P})-2\epsilon)} \max_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n)$$

so

$$2^{n(h(\mathbb{P})-2\epsilon)} \leq \frac{\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n)}{\max_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n)} = \frac{\sum_{z^n \in T_\epsilon^n, \mathbb{Q}(z^n) > 0} \mathbb{Q}(z^n)}{\max_{z^n \in T_\epsilon^n, \mathbb{Q}(z^n) > 0} \mathbb{Q}(z^n)} \leq \#\{z^n \in T_\epsilon^n | \mathbb{Q}(z^n) > 0\}$$

and

$$\log \#\{z^n \in T_\epsilon^n \mid \mathbb{Q}(z^n) > 0\} \geq n(h(\mathbb{P}) - 2\epsilon).$$

So for all $\mathbb{Q} \in \mathcal{B}^{\epsilon,n}(\mathbb{P})$, we have $H_0^n(\mathbb{Q}) > n(h(\mathbb{P}) - 2\epsilon)$. Hence $H_0^{\epsilon,n}(\mathbb{P}) > n(h(\mathbb{P}) - 2\epsilon)$ and $\tilde{h}_0^\epsilon(\mathbb{P}) \geq h(\mathbb{P}) - 2\epsilon$.

□

Theorem 4.4 $\tilde{h}_\infty^\epsilon(\mathbb{P})$ is close to $h(\mathbb{P})$: for all $0 < \epsilon < \frac{1}{2}$:

1. $\tilde{h}_\infty^\epsilon(\mathbb{P}) \geq h(\mathbb{P}) - \epsilon$
2. $\tilde{h}_\infty^\epsilon(\mathbb{P}) \leq h(\mathbb{P}) + 2\epsilon$

Proof Apply theorem 2.27 and let T_ϵ^n be the typical set. For all $z^n \in T_\epsilon^n$, $2^{-n(h(\mathbb{P})+\epsilon)} \leq \mathbb{P}(z^n) \leq 2^{-n(h(\mathbb{P})-\epsilon)}$, and $2^{n(h(\mathbb{P})+\epsilon)} \leq T_\epsilon^n \leq 2^{n(h(\mathbb{P})-\epsilon)}$.

1. Define the distribution

$$\mathbb{P}^{\epsilon,n}(z^n) = \begin{cases} \mathbb{P}(z^n)/\mathbb{P}(T_\epsilon^n) & \text{if } z^n \in T_\epsilon^n, \\ 0 & \text{if } z^n \notin T_\epsilon^n. \end{cases}$$

From comparison of definitions, $F_{\text{norm}}, \mathbb{P}^{\epsilon,n}(z^n) \in \mathcal{B}^{\epsilon,n}(\mathbb{P})$.

$$\tilde{H}_\infty^{\epsilon,n}(\mathbb{P}^{\epsilon,n}) \geq -\log \left(\frac{2^{-n(h(\mathbb{P})-\epsilon)}}{\mathbb{P}(T_\epsilon^n)} \right) = -\log 2^{-n(h(\mathbb{P})-\epsilon)} + \log \mathbb{P}(T_\epsilon^n) = n(h(\mathbb{P})-\epsilon) + \log \mathbb{P}(T_\epsilon^n)$$

So

$$\tilde{h}_\infty^\epsilon(\mathbb{P}) = \lim_{n \rightarrow \infty} \frac{1}{n} \inf_{\mathbb{Q}} \tilde{H}_\infty^n(\mathbb{Q}) \geq \lim_{n \rightarrow \infty} \frac{1}{n} (n(h(\mathbb{P})-\epsilon) + \log \mathbb{P}(T_\epsilon^n)) = \lim_{n \rightarrow \infty} (h(\mathbb{P})-\epsilon) \geq h(\mathbb{P})-\epsilon.$$

2. Let $\mathbb{Q} \in \mathcal{B}^{\epsilon,n}(\mathbb{P})$.

From the definition: $\sum_{z^n} \mathbb{Q}(z^n) \geq 1 - \epsilon$

$$\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n) + \sum_{z^n \notin T_\epsilon^n} \mathbb{Q}(z^n) \geq 1 - \epsilon$$

$$\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n) \geq 1 - 2\epsilon.$$

If n is large enough and $\epsilon < 1/2$, $1 - 2\epsilon \geq 2^{-n\epsilon}$.

Then

$$\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n) \geq 1 - 2\epsilon \geq 2^{-n\epsilon} = 2^{-n(h(\mathbb{P})+2\epsilon)} 2^{n(h(\mathbb{P})+\epsilon)} \geq 2^{-n(h(\mathbb{P})+2\epsilon)} \#T_\epsilon^n$$

so

$$2^{-n(h(\mathbb{P})+2\epsilon)} \leq \frac{\sum_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n)}{\#T_\epsilon^n} \leq \max_{z^n \in T_\epsilon^n} \mathbb{Q}(z^n) \leq \max_{z^n} \mathbb{Q}(z^n)$$

and

$$-\log \max_{z^n} \mathbb{Q}(z^n) \leq n(h(\mathbb{P}) + 2\epsilon).$$

So for all $\mathbb{Q} \in \mathcal{B}^{\epsilon,n}(\mathbb{P})$, we have $H_\infty^n(\mathbb{Q}) \leq n(h(\mathbb{P}) + 2\epsilon)$. Hence $H_\infty^{\epsilon,n}(\mathbb{P}) \leq n(h(\mathbb{P}) + 2\epsilon)$ and $\tilde{h}_\infty^\epsilon(\mathbb{P}) \leq h(\mathbb{P}) + 2\epsilon$.

□

4.4 Quantum case, truncation ball, indirect proof

We prove the quantum case by reduction to the classical case, using the lemmas from section 3.2.

ρ is the distribution being considered. There are n stationary ergodic repetitions $\rho^{(n)}$.

Let λ_i be the eigenvalues of $\rho^{(n)}$.

Define furthermore

$$s(\rho) = \lim_{n \rightarrow \infty} \frac{1}{n} S(\rho^{(n)})$$

$$s_\alpha^\epsilon(\rho) = \lim_{n \rightarrow \infty} \frac{1}{n} S_\alpha^\epsilon(\rho^{(n)})$$

Theorem 4.5 $s_0^\epsilon(\rho)$ is close to $s(\rho)$: for all $0 < \epsilon < \frac{1}{2}$:

1. $s_0^\epsilon(\rho) \leq s(\rho) + \epsilon$
2. $s_0^\epsilon(\rho) \geq s(\rho) - 2\epsilon$

Proof

From the definitions $S(\rho^{(n)}) = H(\underline{\lambda})$, so $s(\rho) = h(\underline{\lambda})$.

Firstly, use lemma 3.7 to get

$$S_0^\epsilon(\rho^{(n)}) = \inf_{\sigma \in \mathcal{B}^\epsilon(\rho^{(n)})} S_0(\sigma) = \inf_{\underline{\mu} \in \mathcal{B}^\epsilon(\underline{\lambda})} H_0(\underline{\mu}) = H_0^\epsilon(\underline{\lambda})$$

Then also

$$\begin{aligned} s_0^\epsilon(\rho) &= \lim_{n \rightarrow \infty} S_0^\epsilon(\rho^{(n)}) \\ &= \lim_{n \rightarrow \infty} H_0^\epsilon(\underline{\lambda}) \\ &= h_0^\epsilon(\underline{\lambda}) \end{aligned}$$

And the classical $h(\underline{\lambda}) - 2\epsilon \leq h_0^\epsilon(\underline{\lambda}) \leq h(\underline{\lambda}) + \epsilon$ gives $s(\rho) - 2\epsilon \leq s_0^\epsilon(\rho) \leq s(\rho) + \epsilon$. \square

Theorem 4.6 $s_\infty^\epsilon(\rho)$ is close to $s(\rho)$: for all $0 < \epsilon < \frac{1}{2}$:

1. $s_\infty^\epsilon(\rho) \geq s(\rho) - \epsilon$
2. $s_\infty^\epsilon(\rho) \leq s(\rho) + 2\epsilon$

Proof

From the definitions $S(\rho^{(n)}) = H(\underline{\lambda})$, so $s(\rho) = h(\underline{\lambda})$.

Firstly, use lemma 3.7 to get

$$S_\infty^\epsilon(\rho^{(n)}) = \inf_{\sigma \in \mathcal{B}^\epsilon(\rho^{(n)})} S_\infty(\sigma) = \inf_{\underline{\mu} \in \mathcal{B}^\epsilon(\underline{\lambda})} H_\infty(\underline{\mu}) = H_\infty^\epsilon(\underline{\lambda})$$

Then also

$$\begin{aligned} s_\infty^\epsilon(\rho) &= \lim_{n \rightarrow \infty} S_\infty^\epsilon(\rho^{(n)}) \\ &= \lim_{n \rightarrow \infty} H_\infty^\epsilon(\underline{\lambda}) \\ &= h_\infty^\epsilon(\underline{\lambda}) \end{aligned}$$

And the classical $h(\underline{\lambda}) - \epsilon \leq h_\infty^\epsilon(\underline{\lambda}) \leq h(\underline{\lambda}) + 2\epsilon$ gives $s(\rho) - \epsilon \leq s_\infty^\epsilon(\rho) \leq s(\rho) + 2\epsilon$. \square

4.5 Quantum case, trace distance ball, indirect proof

The same proof also applies to the trace distance ball, except that the classical statistical distance ball must be used instead of the classical truncation ball.

4.6 Quantum case, truncation ball, direct proof

We also prove the quantum case with the truncation ball without using the lemmas from section 3.2. The proof goes the same way as 4.2.

$\rho^{(n)}$ is the sequence of density matrices being considered; it is stationary and ergodic.

Define furthermore

$$s(\rho) = \lim_{n \rightarrow \infty} \frac{1}{n} S(\rho^{(n)})$$

and

$$s_\alpha^\epsilon(\rho) = \lim_{n \rightarrow \infty} \frac{1}{n} S_\alpha^\epsilon(\rho^{(n)})$$

Theorem 4.7 $s_0^\epsilon(\rho)$ is close to $s(\rho)$: for all $0 < \epsilon < \frac{1}{2}$:

1. $s_0^\epsilon(\rho) \leq s(\rho) + \epsilon$
2. $s_0^\epsilon(\rho) \geq s(\rho) - 2\epsilon$

Proof Use the quantum AEP (theorem 2.30). Let n be an integer such that the AEP holds (i.e. n is greater than or equal to the N given by the AEP) and $1 - 2\epsilon \geq 2^{-n\epsilon}$ and let $\mathcal{T}_\epsilon^{(n)}$ be the corresponding typical subspace.

1. Define the matrix

$$\rho_\epsilon^{(n)} := \rho^{(n)} P_{\mathcal{T}_\epsilon^{(n)}}.$$

Then $0 \leq \rho_\epsilon^{(n)} \leq \rho^{(n)}$ and $\text{tr}(\rho_\epsilon^{(n)}) \geq 1 - \epsilon$, so $\rho_\epsilon^{(n)}$ is in the truncation ball around $\rho^{(n)}$, and $S_0^\epsilon(\rho^{(n)}) \leq S_0(\rho_\epsilon^{(n)})$.

From the definition of Rényi entropy:

$$S_0(\rho_\epsilon^{(n)}) = \log \text{rank}(\rho_\epsilon^{(n)})$$

Because the rank of a matrix product cannot be more than the rank of a factor:

$$\log \text{rank}(\rho_\epsilon^{(n)}) \leq \log \text{rank}(P_{\mathcal{T}_\epsilon^{(n)}})$$

Because $P_{\mathcal{T}_\epsilon^{(n)}}$ is a projection with eigenvalues 0 and 1 only (theorem 2.4):

$$\log \text{rank}(P_{\mathcal{T}_\epsilon^{(n)}}) = \log \text{tr}(P_{\mathcal{T}_\epsilon^{(n)}})$$

From the properties of the typical subspace,

$$\log \text{tr}(P_{\mathcal{T}_\epsilon^{(n)}}) \leq \log 2^{n(s(\rho) + \epsilon)}.$$

So

$$S_0^\epsilon(\rho^{(n)}) \leq n(s(\rho) + \epsilon).$$

and

$$s_0^\epsilon(\rho) = \lim_{n \rightarrow \infty} \frac{1}{n} S_0^\epsilon(\rho^{(n)}) \leq s(\rho) + \epsilon.$$

2. Let $\sigma^{(n)} \in \mathcal{B}_n^\epsilon(\rho^{(n)})$.

From the definition: $\text{tr}(\sigma^{(n)}) \geq 1 - \epsilon$. Then $\text{tr}(\sigma^{(n)} P_{\mathcal{T}_\epsilon^{(n)}}) + \text{tr}(\sigma^{(n)}(I - P_{\mathcal{T}_\epsilon^{(n)}})) \geq 1 - \epsilon$.

Note that $\text{tr}(\rho^{(n)}(I - P_{\mathcal{T}_\epsilon^{(n)}})) \leq \epsilon$. Use $\rho^{(n)} - \sigma^{(n)} \geq 0$ with the fact that $I - P_{\mathcal{T}_\epsilon^{(n)}}$ is a projection, then $(\rho^{(n)} - \sigma^{(n)})(I - P_{\mathcal{T}_\epsilon^{(n)}}) \geq 0$.

So $\text{tr}(\sigma^{(n)}(I - P_{\mathcal{T}^{(n)}})) \leq \epsilon$ and

$$\text{tr}(\sigma^{(n)}P_{\mathcal{T}^{(n)}}) \geq 1 - 2\epsilon.$$

If n is large enough and $\epsilon < 1/2$, $1 - 2\epsilon \geq 2^{-n\epsilon}$.

Then

$$\begin{aligned} \text{tr}(\sigma^{(n)}P_{\mathcal{T}^{(n)}}) &\geq 1 - 2\epsilon \geq 2^{-n\epsilon} = 2^{n(s(\rho)-2\epsilon)}2^{-n(s(\rho)-\epsilon)} \\ &\geq 2^{n(s(\rho)-2\epsilon)}\lambda_{\max}(\rho^{(n)}P_{\mathcal{T}^{(n)}}) \geq 2^{n(s(\rho)-2\epsilon)}\lambda_{\max}(\sigma^{(n)}P_{\mathcal{T}^{(n)}}) \end{aligned}$$

so

$$\begin{aligned} 2^{n(s(\rho)-2\epsilon)} &\leq \frac{\text{tr}(\sigma^{(n)}P_{\mathcal{T}^{(n)}})}{\lambda_{\max}(\sigma^{(n)}P_{\mathcal{T}^{(n)}})} \\ &\leq \frac{\lambda_{\max}(\sigma^{(n)}P_{\mathcal{T}^{(n)}})\text{rank}(\sigma^{(n)}P_{\mathcal{T}^{(n)}})}{\lambda_{\max}(\sigma^{(n)}P_{\mathcal{T}^{(n)}})} = \text{rank}(\sigma^{(n)}P_{\mathcal{T}^{(n)}}) \leq \text{rank}(\sigma^{(n)}) \end{aligned}$$

and

$$S_0(\sigma^{(n)}) = \log \text{rank}(\sigma^{(n)}) \geq n(s(\rho) - 2\epsilon).$$

From the definition of smooth Rényi entropy,

$$S_0^\epsilon(\rho^{(n)}) \geq n(s(\rho) - 2\epsilon)$$

and in the limit

$$s_0^\epsilon(\rho) = \lim_{n \rightarrow \infty} \frac{1}{n} S_0^\epsilon(\rho^{(n)}) \geq s(\rho) - 2\epsilon.$$

□

Theorem 4.8 $s_\infty^\epsilon(\rho)$ is close to $s(\rho)$: for all $0 < \epsilon < \frac{1}{2}$:

1. $s_\infty^\epsilon(\rho) \geq s(\rho) - \epsilon$
2. $s_\infty^\epsilon(\rho) \leq s(\rho) + 2\epsilon$

Proof Use the quantum AEP (theorem 2.30). Let n be an integer such that the AEP holds (i.e. n is greater than or equal to the N given by the AEP) and $1 - 2\epsilon \geq 2^{-n\epsilon}$ and let $\mathcal{T}_\epsilon^{(n)}$ be the corresponding typical subspace.

1. Define the matrix

$$\rho_\epsilon^{(n)} := \rho^{(n)}P_{\mathcal{T}_\epsilon^{(n)}}.$$

Then $0 \leq \rho_\epsilon^{(n)} \leq \rho^{(n)}$ and $\text{tr}(\rho_\epsilon^{(n)}) \geq 1 - \epsilon$, so $\rho_\epsilon^{(n)}$ is in the truncation ball around $\rho^{(n)}$, and $S_\infty^\epsilon(\rho^{(n)}) \geq S_\infty^\epsilon(\rho_\epsilon^{(n)})$.

From the definition of Rényi entropy and the properties of the typical subspace,

$$S_\infty(\rho_\epsilon^{(n)}) = -\log \lambda_{\max}(\rho_\epsilon^{(n)}) \geq -\log 2^{-n(s(\rho)-\epsilon)} = n(s(\rho) - \epsilon).$$

So

$$S_\infty^\epsilon(\rho^{(n)}) \geq n(s(\rho) - \epsilon).$$

and

$$s_\infty^\epsilon(\rho) = \lim_{n \rightarrow \infty} \frac{1}{n} S_\infty^\epsilon(\rho^{(n)}) \geq s(\rho) - \epsilon.$$

2. Let $\sigma^{(n)} \in \mathcal{B}_n^\epsilon(\rho^{(n)})$.

From the definition: $\text{tr}(\sigma^{(n)}) \geq 1 - \epsilon$. Then $\text{tr}(\sigma^{(n)} P_{\mathcal{T}_\epsilon^{(n)}}) + \text{tr}(\sigma^{(n)}(I - P_{\mathcal{T}_\epsilon^{(n)}})) \geq 1 - \epsilon$.

Note that $\text{tr}(\rho^{(n)}(I - P_{\mathcal{T}_\epsilon^{(n)}})) \leq \epsilon$. Use $\rho^{(n)} - \sigma^{(n)} \geq 0$ with the fact that $I - P_{\mathcal{T}_\epsilon^{(n)}}$ is a projection, then $(\rho^{(n)} - \sigma^{(n)})(I - P_{\mathcal{T}_\epsilon^{(n)}}) \geq 0$.

So $\text{tr}(\sigma^{(n)}(I - P_{\mathcal{T}_\epsilon^{(n)}})) \leq \epsilon$ and

$$\text{tr}(\sigma^{(n)} P_{\mathcal{T}_\epsilon^{(n)}}) \geq 1 - 2\epsilon.$$

If n is large enough and $\epsilon < 1/2$, $1 - 2\epsilon \geq 2^{-n\epsilon}$.

Then

$$\text{tr}(\sigma^{(n)} P_{\mathcal{T}_\epsilon^{(n)}}) \geq 1 - 2\epsilon \geq 2^{-n\epsilon} = 2^{-n(s(\rho)+2\epsilon)} 2^{n(s(\rho)+\epsilon)} \geq 2^{-n(s(\rho)+2\epsilon)} \text{tr}(P_{\mathcal{T}_\epsilon^{(n)}})$$

so

$$2^{-n(s(\rho)+2\epsilon)} \leq \frac{\text{tr}(\sigma^{(n)} P_{\mathcal{T}_\epsilon^{(n)}})}{\text{tr}(P_{\mathcal{T}_\epsilon^{(n)}})} \leq \lambda_{\max}(\sigma^{(n)} P_{\mathcal{T}_\epsilon^{(n)}}) \leq \lambda_{\max}(\sigma^{(n)}).$$

For the second inequality, note that (replacing all eigenvalues with the largest eigenvalue)

$$\sigma P_{\mathcal{T}_\epsilon^{(n)}} \leq \lambda_{\max}(\sigma P_{\mathcal{T}_\epsilon^{(n)}}) I$$

and (multiplying both sides by P and taking the trace)

$$\text{tr}(\sigma P_{\mathcal{T}_\epsilon^{(n)}}^2) \leq \text{tr}(\lambda_{\max}(\sigma P_{\mathcal{T}_\epsilon^{(n)}}) P_{\mathcal{T}_\epsilon^{(n)}}) = \lambda_{\max}(\sigma P_{\mathcal{T}_\epsilon^{(n)}}) \text{tr}(P_{\mathcal{T}_\epsilon^{(n)}}).$$

For the third inequality, note that $P_{\mathcal{T}_\epsilon^{(n)}}$ is a projection, $\lambda_{\max}(\sigma^{(n)} P_{\mathcal{T}_\epsilon^{(n)}}) \leq \lambda_{\max}(\sigma^{(n)})$.

Then

$$S_\infty(\sigma^{(n)}) = -\log \lambda_{\max}(\sigma^{(n)}) \leq -n(s(\rho) + 2\epsilon).$$

From the definition of smooth Rényi entropy,

$$S_{\infty}^{\epsilon}(\rho^{(n)}) \leq n(s(\rho) + 2\epsilon)$$

and in the limit

$$s_{\infty}^{\epsilon}(\rho) = \lim_{n \rightarrow \infty} \frac{1}{n} S_{\infty}^{\epsilon}(\rho^{(n)}) \leq s(\rho) + 2\epsilon.$$

□

Chapter 5

Conclusions

In the limit of the number of repetitions $n \rightarrow \infty$ and $\epsilon \rightarrow 0$, smooth Rényi entropy is equal to Shannon entropy, for stationary ergodic sources, both of classical and quantum information.

This means it is possible to use Shannon entropy for asymptotic security proofs of cryptographic protocols which naturally call for smooth Rényi entropy, also in the quantum ergodic case. Examples of such cryptographic protocols are information-theoretically secure key exchange protocols.

The truncation ball is easier to work with than the statistical distance or trace distance ball, as it is not necessary to normalize elements. Although there is a difference between smooth Rényi entropy using the truncation ball and smooth Rényi entropy using the statistical distance or trace distance ball, this vanishes in the limit for $\epsilon \rightarrow 0$. Hence, they are effectively the same and the truncation ball is preferable. There may be other balls with even better properties.

The proofs do not use ergodicity directly, only the AEP. This means that the theorems also hold for more general information sources which are not stationary ergodic but do have the AEP property.

The constant 2 in $h_0^\epsilon(\mathbb{P}) \geq h(\mathbb{P}) - 2\epsilon$ in theorem 4.1 and in $h_\infty^\epsilon(\mathbb{P}) \leq h(\mathbb{P}) + 2\epsilon$ in theorem 4.2 can be replaced by any $1 + \delta$ with $\delta > 0$, and also in the other analogous theorems in chapter 4.

Our theorems do not apply to the conditional case. A good suggestion for further research would be to extend them to this case. There are complications in the quantum case; the direct proof for the quantum case (section 4.6) is a good starting point.

Another suggestion for further research is infinite alphabets (sets \mathcal{Z}), which can bring the definitions based on the two balls closer together and allow more generality, but cause various complications.

Bibliography

- [1] R. Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, 1997.
- [2] Igor Bjelaković and Arleta Szkola. The data compression theorem for ergodic quantum information sources, January 2003.
- [3] Christian Cachin. Smooth entropy and Rényi entropy. *Lecture Notes in Computer Science*, 1233:193–208, 1997.
- [4] Thomas N. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley and Sons, Inc., 1991.
- [5] Michael E. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2001.
- [6] R. Renner and R. König. Universally composable privacy amplification against quantum adversaries. In *TCC 2005, LNCS 3378*, 2005. Also available as [quant-ph/0403133](http://arxiv.org/abs/quant-ph/0403133).
- [7] R. Renner and S. Wolf. Smooth Rényi entropy and applications. In *ISIT*, 2004. Full version available as <http://qi.ethz.ch/pub/publications/smooth.ps>.
- [8] R. Renner and S. Wolf. Simple and tight bounds for information reconciliation and privacy amplification. In *ASIACRYPT*, 2005.
- [9] Stefan Wolf. Unconditional security in cryptography. In Ivan Damgård, editor, *Lectures on data security: modern cryptology in theory and practise*, volume 1561 of *Lecture Notes in Computer Science*, pages 217–250. Springer-Verlag, July 1998.