Radboud University

Master Thesis in Mathematics

Track: Mathematical Physics

A Universal Quantum Sanov Theorem

Jacques Janse van Rensburg

August, 2026

Supervisor: Klaas Landsman
Second reader: Manus Visser

Abstract

In this thesis we investigate the possibility of extending Sanov's theorem in large deviation theory to the quantum setting. A complete proof of the classical Sanov theorem and the closely related Stein lemma is provided. Additionally, the quantum Stein lemma is proved and we provide an overview of two proposed quantum Sanov theorems in literature. We especially focus on the role of the Umegaki relative entropy appearing in the quantum Sanov theorem proposed by Bjelaković et al. (CMP, 2005). Our first result is a no-go theorem that proves in sufficiently large Hilbert spaces ($\dim(\mathcal{H}) \geq 4$), it is impossible to construct a sequence of tests that separates a set of normal states $\mathcal{A}$, from arbitrary normal states σ at a rate given by the Umegaki relative entropy. A new quantum extension of Sanov's Theorem is proved, which we call a universal quantum Sanov theorem. For any open set of states $\mathcal{A}$, we construct a sequence of tests that can separate any state $\sigma \notin \mathcal{A}$ from $\mathcal{A}$ at an exponential rate. To our knowledge, such a universality result has not previously been established in the quantum regime, and it offers a new perspective on the concept of a quantum empirical measure. Using von Neumann algebras, we finish this thesis by extending our quantum Sanov theorem to a specific infinite dimensional non-commutative setting.

Contents

I Introduction **5**
- I.1 Outline of thesis . . . 7
- I.2 Acknowledgements . . . 7

II Classical probability **9**
- II.1 Preliminaries . . . 10
- II.2 Finite spaces . . . 13
  - II.2.1 Hypothesis testing . . . 14
  - II.2.2 Sanov's theorem . . . 20
  - II.2.3 Implications of Sanov's theorem . . . 27
- II.3 Sanov's theorem in more generality . . . 39

III Quantum probability **41**
- III.1 Preliminaries . . . 42
- III.2 The quantum formalism of classical probability . . . 46
- III.3 Quantum measurement and hypothesis testing . . . 49
  - III.3.1 Measuring quantum states . . . 49
  - III.3.2 Quantum hypothesis testing . . . 52
  - III.3.3 Proof of the quantum Stein lemma . . . 55
- III.4 The first quantum Sanov theorem . . . 59
- III.5 Another quantum Sanov theorem . . . 61

IV A universal quantum Sanov theorem for type I_d **64**
- IV.1 Universality . . . 65
- IV.2 Universal quantum Sanov theorem . . . 69
  - IV.2.1 An example . . . 73

V A universal quantum Sanov theorem in more generality **75**
- V.1 Preliminaries . . . 76
- V.2 Information complete operator valued measures . . . 77
  - V.2.1 Constructed from predual . . . 77
  - V.2.2 Constructed from a norm separable proper ideal . . . 80
- V.3 Extending the universal Sanov theorem . . . 82
  - V.3.1 Type I_∞ . . . 82
  - V.3.2 A counter example constructed by Vaes . . . 85

VI Conclusion and outlook **88**

Bibliography **89**

A Selected topics in mathematics **93**
A.1 Nets in topology 93
A.2 Probability theory 95
A.3 Schur-Weyl Duality 98
A.4 Miscellaneous results 101

B Operator algebras **104**

Chapter I

Introduction

Randomness is an intrinsic part of quantum mechanics. From the perspective of this work, quantum mechanics could just as well be called a theory of non-commutative probability, although the latter is more colloquially referred to as quantum probability. In the field of classical probability, as formalized by Kolmogorov, one uses the concepts of σ-algebras, measurable functions, and probability measures to describe random behaviour. On the other hand, the mathematical theory of quantum mechanics, as formalized by von Neumann, uses the language of von Neumann algebras, self adjoint operators, and positive unital normal functionals to describe randomness. The mathematical framework of quantum mechanics laid out by von Neumann and its probabilistic interpretation as pioneered by Born, is the foundation of quantum probability theory. It is worthwhile to point out that in the entirety of this text probability distributions and quantum states are the essential connection between mathematics and randomness. We distinguish classical from quantum randomness purely from the mathematical language one is required to use to describe the underlying system. We stress the fact that classical probability theory is an insufficient tool to describe all random systems in our universe, and quantum probability is the mathematical remedy. [1]

The perspective that quantum probability is a non-commutative version of classical probability is supported by the fact that every commutative von Neumann algebra $\mathcal{M}$ is isomorphic to a space of essentially bounded measurable functions $L^\infty(X, \mu)$, for some locally compact space X and probability measure μ on X. Essentially, if the observables describing some *quantum* system commute, then the system can be described classically. At the heart of this thesis lies the perspective that while quantum mechanics is non-classical, it is crucial that we extend notions and results in classical probability to the quantum setting. We believe it is through this process that the distinguishing features of classical and quantum randomness can be best understood. Sticking to this narrative, this thesis largely focuses on proving several quantum extensions of the Sanov theorem in classical probability. Proven in 1957 by Ivan Nikolaevich Sanov, this result relates the Kullback-Leibler relative entropy to the probability that the empirical observations obtained from repeatedly sampling a random variable on some probability space X with law p, appears as if X has law $q \in \mathcal{A} \subset \text{Prob}(X)$ for some measurable set $\mathcal{A}$.

As a first goal of this work, we will give an overview of the Sanov theorem in its classical

[1] We direct the interested reader to [25] for an overview of the interpretations of randomness in mathematics and physics.

glory along with the existing extensions of the Sanov theorem to the quantum setting. We will follow an identical path for the closely related Stein lemma, which has also been extended to the quantum regime. As we will see the theory of (quantum) hypothesis testing, and many forms of relative entropy, are essential tools for describing the optimal rate at which one can separate an unknown prior state from a collection of other states. After the literature review the rest of the thesis is dedicated to presenting the new results obtained during this internship. In the spirit of using classical theory as the foundation of understanding quantum systems, this involves reducing a quantum system to a classical one while preserving certain structures, which in our case is the distinguishability of quantum states. A crucial part of reducing the states in a quantum setting $\mathcal{S}(\mathcal{M})$ to classical probability distributions $\Lambda \subseteq \text{Prob}(X)$ is making use of an information complete positive operator valued measure $\mu : \Omega \to \mathcal{M}^+$, where Ω is a σ-algebra of the underlying space X. Moreover, the information complete positive operator valued measure μ can then be used to lift results from the classical setting into the quantum setting. With this technique we introduce a new variant of the quantum Sanov theorem, which we call a *universal* quantum Sanov theorem. Our result is a direct analogue to the classical Sanov theorem and goes beyond the usual finite dimensional setting (type I_d) that research in this field is often restricted to.

I.1 Outline of thesis

In Chapter II we begin with a preliminary section discussing classical probability. We mostly restrict ourselves to studying probability distributions over finite sets, and introduce hypothesis testing in this context. Thereafter, we provide a proof of Sanov's theorem, and in Section II.2.3 we demonstrate how Sanov's theorem can be used to prove fundamental results such as Stein's lemma, Chernoff's lemma, and the strong law of large numbers. The last section of this chapter provides an overview of Sanov's theorem for distributions over Polish spaces and measure spaces.

In Chapter III we follow a similar structure as in Chapter II, however, we instead focus on quantum probability theory. To illustrate how quantum probability can be viewed as a non-commutative extension of classical probability, in Section III.2 we restate the classical Sanov theorem in the setting of (commutative) quantum probability. In Section III.3.1 we introduce the mathematical framework of measuring quantum states. We provide a proof of the quantum Stein lemma, which will be used to sketch a proof of one of two quantum Sanov theorem's we investigate from literature.

The rest of this thesis is dedicated to showcasing our own research. In Chapter IV we propose a new quantum Sanov theorem, that we call a universal quantum Sanov theorem. In Section IV.1 we discuss the notion of universality, and prove in the quantum setting it is not possible to have a test that separates a collection of normal states $\mathcal{A}$ from an arbitrary state $\sigma \notin \mathcal{A}$ at a rate given by the Umegaki relative entropy. This result crucially relies on using a pair of non-commuting quantum states, hence this result does not hold in the classical setting. The new quantum Sanov theorem we propose is stated and proved in Section IV.2, and we also propose a quantum extension of the classical empirical measure.

In Chapter V we extend our universal quantum Sanov theorem to an infinite dimensional setting using von Neumann algebras. Finally, in Chapter VI we conclude our work and give an outlook of several possible new research directions.

I.2 Acknowledgements

I am grateful to Ludovico Lami and Bartosz Reguła for meeting with me to discuss the concept of universal state separation and sharing relevant literature. Most notably they brought it to my attention that a result in [28] is very relevant to my research. I would like to thank Martijn Casper for taking time to discuss a critical unresolved question I was faced with. And additionally, I would like to thank Stefaan Vaes, who was able to find a solution to my question (detailed in Section V.3). Finally, I would like to thank my supervisor Klaas Landsman for all the support and guidance provided throughout my internship.

Data management

☒ This thesis is based entirely on existing data, literature, and theory.

☒ No new research data was collected or generated.

☒ No data sets were stored or processed as part of this thesis.

☒ No personal data was processed.

No data were generated or used for the preparation of this thesis.

Use of generative AI

☒ This thesis was written by the author and no generative AI was used to produce text or proofs.

☒ Services such as Google AI overview, ChatGPT, and Claude AI were used to search literature for specific mathematical statements and concepts, and used for checking spelling/grammar.

Chapter II

Classical probability

The primary goal of this chapter is to introduce, motivate, and prove Sanov's theorem and Stein's lemma. As alluded in the introduction, an operational interpretation of the Sanov theorem is that it describes the effectiveness of using empirical observations as a basis for deciding if a probability measure p belongs to some set $\mathcal{A} \subset \text{Prob}(X)$. We wish to formalize this interpretation through the lens of hypothesis testing, which as we will soon appreciate has very strong connections to Stein's lemma. Although, to an even greater extent we hope to illustrate that the Sanov theorem provides more than this helpful, but limited, operational interpretation. For the rest of this section we assume the reader is familiar with basic measure theory and probability theory. Nonetheless, we include a short section in the appendix listing all essential definitions and results in measure theoretic probability theory. As a hopefully rare deviation from our streamlined approach, we now recall the axioms of probability as stated by Kolmogorov in his book *Foundations of the Theory of Probability* published in 1933. Put briefly, let X be a set and Ω a σ-algebra (Definition A.2.1) of subsets of X. Then a non-negative function $p : \Omega \to \mathbb{R}_{0\geq}$ assigns a probability to each set $A \in \Omega$, called an event if,

1. $p(X) = 1$,
2. for all $A \in \mathcal{F}$, $p(A) \in [0, 1]$, where $p(A)$ is interpreted as the probability of the event A occurring,
3. for a sequence of disjoint sets $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F}$ we have
$$p\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} p(A_n).$$

This list of axioms is found in virtually any introductory treatment of probability theory. Our motivation for stating them here is two fold. Firstly, we find it very insightful to compare the classical foundations of probability to the much more subtle foundations in quantum probability. And secondly, we believe Kolmogorov's axioms of probability are rooted in the necessarily empirical interactions one has with randomness in nature. In a very complementary manner, Sanov's theorem connects the empirical observations we can make in a random system with the beautiful concept of entropy originating in statistical physics. The rest of this chapter is continued as follows. In addition to providing a proof of Sanov's theorem we will also give a complete account of the closely associated Stein's Lemma. First, these two results will be treated for finite probability spaces, and then we will move on to the more general Polish space setting. We begin with a preliminary section.

II.1 Preliminaries

No results in this section will be proved. We refer the reader to [12] for a more detailed treatment of this material.

Definition II.1.1. *A topological space (X, τ) is **Polish** if it is separable, metrizable by some metric d, and complete with respect to d. Unless stated otherwise, we always consider X as a measure space $(X, \mathcal{F})$, where $\mathcal{F}$ is the Borel σ-algebra generated by the open sets τ, we call the pair $(X, \mathcal{F})$ a **standard Borel space**. Moreover, with* $\mathrm{Prob}(X)$ *we denote the set of all Borel probability measures over X.*

There are several topologies one can define on the set $\mathrm{Prob}(X)$, the most useful one at our disposable is called the weak topology.

Definition II.1.2 (Weak topology on $\mathrm{Prob}(X)$)**.** *The **weak topology** on* $\mathrm{Prob}(X)$ *is generated by the basis of open sets*

$$U_{f,x,\delta} := \Big\{ p \in \mathrm{Prob}(X); \Big| \int_X f \mathrm{d}p - x \Big| < \delta \Big\},$$

where $x \in \mathbb{R}$, $\delta > 0$ and $f \in C_b(X)$ is a continuous and bounded function. Equivalently, a net of probability measures $(p_i)_{i \in I}$ converges to a probability measure $q \in \mathrm{Prob}(X)$ if for all $f \in C_b(X)$ we have

$$\lim_i \int_X f(x) \mathrm{d}p_i(x) = \int_X f(x) \mathrm{d}q. \tag{2.1}$$

A brief overview of the role of nets in topology is given in the Appendix A.1. However, the necessity of using nets instead of sequences will be short lived in this chapter, as the weak topology is in fact metrizable by the Lévy metric when X is Polish.

Definition II.1.3 (Lévy metric)**.** *Let (X, d) be Polish and suppose $A \subseteq X$ is a measurable set, then for $\delta > 0$ define*

$$A^\delta := \{x \in X; \exists a \in A \text{ such that } d(x, a) < \delta\}.$$

*The **Lévy metric** $d_L : \mathrm{Prob}(X) \times \mathrm{Prob}(X) \to \mathbb{R}_{0\geq}$ is then defined as,*

$$d_L(p, q) := \inf\{\delta > 0; p(A) < q(A^\delta) + \delta \text{ for all closed } A \subseteq X\}. \tag{2.2}$$

Note that for all measurable $A \subseteq X$ and $\delta > 0$ the set A^δ is measurable. Since we assumed X is Polish, it is separable and there is a countable dense subset $(a_n)_{n \in \mathbb{N}}$ of A. Then A^δ is the countable union of open balls in X,

$$A^\delta = \bigcup_{n \in \mathbb{N}} \{x \in X; d(x, a_n) < \delta\}.$$

Theorem II.1.1. *Assume X is a Polish space. Then we have the following:*

1. *The Lévy metric generates the weak topology on Prob(X).*
2. *The space Prob(X) is also Polish when equipped with d_L.*
3. *Prob(X) is weakly compact if and only if X is compact.*

Proof. See [12, Theorem D.8]. □

We briefly want to discuss another relevant topology one can apply on the space of probability measures $\text{Prob}(X)$ called the $\tau-$topology, which is a stronger topology than the weak topology on $\text{Prob}(X)$. Note that we implicitly defined the weak topology on $\text{Prob}(X)$ only for X Polish, but this is only done since the weak topology is then compatible with the Lévy metric. It will be useful to define the τ-topology for the more general setting that X is a measure space.

Definition II.1.4 (τ-topology on $\text{Prob}(X)$)**.** *Let X be a measure space. Then the* ***τ-topology*** *on $\text{Prob}(X)$ is generated by the basis of open sets*

$$W_{g,x,\delta} := \Big\{ p \in \text{Prob}(X); \Big| \int_X g \mathrm{d}p - x \Big| < \delta \Big\},$$

where $x \in \mathbb{R}$, $\delta > 0$ and $g \in L^\infty(X)$ is an essentially bounded Borel measurable function.

If a net $(p_i)_{i\in I}$ converges to some $q \in \text{Prob}(X)$ in the $\tau-$topology, that is, for all $f \in L^\infty(X)$,

$$\lim_i \int_X f \mathrm{d}p_i = \int_X f \mathrm{d}q,$$

then the net also converges in the weak topology since $C_b(X) \subseteq L^\infty(X)$. The τ-topology will only return towards the end of this chapter when we discuss even further extensions of the Sanov theorem from Polish spaces to general measure spaces. We round off this preliminary section by defining the three essential concepts, the empirical measure L_n, and two types of relative entropy entropies, one being the Kullback-Leibler relative entropy $D_{KL}(\cdot||\cdot)$ and the other the α-Rényi relative entropy $R_\alpha(\cdot||\cdot)$ for $\alpha \in \mathbb{R}\backslash\{0,1\}$.

Definition II.1.5 (Empirical Measure)**.** *Let X be a set and $x_1 \in X$, then the* ***Dirac measure*** *at x_1, denoted as $\delta_{x_1} \in \text{Prob}(X)$, is defined as*

$$\delta_{x_1}(A) := \begin{cases} 1 & \text{if } x_1 \in A \\ 0 & \text{if } x_1 \notin A. \end{cases}$$

For $n \in \mathbb{N}$ the function $L_n : X^{\mathbb{N}} \to \text{Prob}(X)$ is called the ***empirical measure*** *and is defined as*

$$L_n(x) := \frac{1}{n} \sum_{k=1}^{n} \delta_{x_k}, \tag{2.3}$$

$x = (x_1, \cdots, x_n, \cdots) \in X^{\mathbb{N}}$.

We note that for $n \in \mathbb{N}$ the function L_n depends only on the first n coordinates of $x \in X^{\mathbb{N}}$, hence we slightly abuse of notation and often consider L_n as a function from the n-fold Cartesian product of X, namely $L_n : A^n \to \text{Prob}(X)$. An important conceptual remark about the empirical measure is the following. Given some probability distribution $p \in \text{Prob}(X)$ one can take n *samples* of identical random processes each with probability distribution p, producing a string of observations $x = (x_1, \cdots, x_n) \in X^n$. Based on the observational data x, the empirical distribution $L_n(x)$ is the frequency that each observation occurred in the string $(x_1, \cdots, x_N)$. This precisely aligns with the interpretation

of probability theory proposed by Kolmogorov, namely for large enough n the frequency of the event x_i occurring in the string x will approach the probability $p(x_i)$ of that event occurring. In an intrinsic way, the asymptotic behaviour of the empirical measure is therefore at the heart of probability. One indispensable tool to describe this asymptotic behaviour is the notion of entropy, and in particular the Kullback-Leibler relative entropy.

Definition II.1.6 (Kullback-Leibler relative entropy)**.** *Let $(X, \mathcal{F})$ be a measure space and $p, q \in \mathrm{Prob}(X)$. Then we say p is **absolutely continuous** with respect to q and write $p \ll q$ if for all $A \in \mathcal{F}$ we have $q(A) = 0 \implies p(A) = 0$. The **Kullback-Leibler relative entropy** $D_{KL}(\cdot||\cdot) : \mathrm{Prob}(X) \times \mathrm{Prob}(X) \to [0, \infty]$ is defined as*

$$D_{KL}(p||q) := \begin{cases} \int_X \frac{\mathrm{d}p}{\mathrm{d}q} \log\left(\frac{\mathrm{d}p}{\mathrm{d}q}\right) \mathrm{d}q & p \ll q, \\ \infty & \text{otherwise,} \end{cases} \tag{2.4}$$

where $\frac{\mathrm{d}p}{\mathrm{d}q} : X \to \mathbb{R}$ is the Radon-nikodym derivative (Theorem A.2.2).

Another entropy we will come across is the Rényi relative entropy.

Definition II.1.7 (Rényi Relative Entropy)**.** *Let $(X, \mathcal{F})$ be a measure space and $p, q \in \mathrm{Prob}(X)$. Take any probability measure $\mu \in \mathrm{Prob}(A)$ such that p and q are absolutely continuous with respect to μ. For $\alpha \in \mathbb{R}\backslash\{0, 1\}$ the α-**Rényi relative entropy** $R_\alpha(\cdot||\cdot) : \mathrm{Prob}(X) \times \mathrm{Prob}(X) \to [0, \infty)$ is defined as*

$$R_\alpha(p||q) := \frac{1}{\alpha - 1} \log \int_X \left(\frac{\mathrm{d}p}{\mathrm{d}\mu}\right)^\alpha \left(\frac{\mathrm{d}q}{\mathrm{d}\mu}\right)^{1-\alpha} \mathrm{d}\mu. \tag{2.5}$$

where $\frac{\mathrm{d}p}{\mathrm{d}\mu}$ and $\frac{\mathrm{d}q}{\mathrm{d}\mu}$ are the Radon-nikodym derivatives with respect to μ. We remark that the Rényi relative entropy is independent of the measure μ.

Finally, we state a collection of useful and well known properties of the Kullback-Leibler and Rènyi relative entropy.

Proposition II.1.1. *Let $(X, \mathcal{F})$ be a measure space and $p, q \in \mathrm{Prob}(X)$.*

1. *$D_{KL}(p||q) \geq 0$ and $D_{KL}(p||q) = 0$ if and only if $p = q$.*
2. *The map $p, q \mapsto D_{KL}(p||q)$ is jointly convex.*
3. *If X is Polish, then the map $p, q \mapsto D_{KL}(p||q)$ is lower semicontinuous (Definition A.2.2) with respect to the product weak topology.*

Proof. See [33, Lemma 5.3, Theorem 5.4, Theorem 5.6]. □

Proposition II.1.2. *Let $(X, \mathcal{F})$ be a measure space and $p, q \in \mathrm{Prob}(X)$. Then for all $\alpha \in \mathbb{R}\backslash\{0, 1\}$,*

1. *$R_\alpha(p||q) \geq 0$.*
2. *The map $p, q \mapsto R_\alpha(p||q)$ is jointly convex.*
3. *If X is Polish, the map $p, q \mapsto R_\alpha(p||q)$ is lower semicontinuous with respect to the product weak topology.*

Proof. See [36]. □

II.2 Finite spaces

Let A be a finite non empty set of cardinality $|A| = d$. We can now interpret $\text{Prob}(A) \subseteq \mathbb{R}^d$ as the simplex

$$\Big\{(x_1, \cdots, x_d) \in \mathbb{R}^d; \sum_{i=1}^{d} x_i = 1, x_i \geq 0 \text{ for all } i = 1, \cdots, n\Big\}, \tag{2.6}$$

of $\mathbb{R}^d$. Naturally the weak topology, τ-topology, and Euclidean topology all coincide in the finite dimensional setting. If $p, q \in \text{Prob}(A)$ then p is absolutely continuous with respect to q iff for all $a \in A$, $q(a) = 0 \implies p(a) = 0$. Additionally, the Kullback-Leibler and Rényi relative entropies respectively reduce to

$$D_{KL}(p||q) = \begin{cases} \sum_{a\in A} p(a)(\log p(a) - \log q(a)) & p \ll q, \\ \infty & \text{otherwise}, \end{cases} \tag{2.7}$$

where we use the convention that $0 \cdot \log 0 = 0$, and,

$$R_\alpha(p||q) = \frac{1}{\alpha - 1} \log \sum_{a\in A} (p(a))^\alpha (q(a))^{1-\alpha}. \tag{2.8}$$

One of the advantages of starting our discussion in the finite dimensional setting is that the interpretation of probability theory is in many ways intuitive. We start of with the very elementary formalism of what the probability of an event occurring means mathematically. For a probability $p \in \text{Prob}(A)$ and $n \in \mathbb{N}$ one can extend p to its Bernoulli measure $p^n \in \text{Prob}(A^n)$ by specifying

$$p^n(a_1, \cdots, a_n) = p(a_1) \cdots p(a_n). \tag{2.9}$$

The Bernoulli measure p^n is the probability measure that describes the process of sampling n independent and identically distributed (i.i.d) random processes that are governed by the probability measure p. By the Carathéodory extension theorem (Theorem A.2.1) there exists a unique probability measure $p^{\mathbb{N}}$ on the measure space $(A^{\mathbb{N}}, \mathcal{F})$, where $\mathcal{F}$ is the σ-algebra generated by cylinder sets

$$[a_1, \cdots, a_n] := \{b \in A^{\mathbb{N}}; a_i = b_i \text{ for } i = 1, \cdots, n\}, \tag{2.10}$$

such that

$$p^{\mathbb{N}}([a_1, \cdots, a_n]) = p^n(a_1, \cdots, a_n), \tag{2.11}$$

for $n \in \mathbb{N}$ and $(a_1, \cdots, a_n) \in A^n$. Then we have the following foundational result.

Proposition II.2.1. *Let A be a finite set and $p \in \text{Prob}(A)$. Then for almost every $\sigma \in A^{\mathbb{N}}$,*

$$\lim_{N\to\infty} L_n(\sigma)(a) = p(a), \quad \forall a \in A. \tag{2.12}$$

Or equivalently L_n converges weakly to p in $\text{Prob}(A)$ as $n \to \infty$.

Proposition II.2.1 supports the underlying principle that the probability $p(a)$ of an event $a \in A$ is the frequency one observes a upon repeatedly sampling a random variable on A with law p. This proposition is our first of many results which describe the asymptotic behaviour of the empirical measure. We postpone the proof of Proposition II.2.1 since we wish to show it is a consequence of Sanov's theorem in Section II.2.3. We closely follow parts of [26] in this argument. We should mention that our approach is by no means necessary, since one can instead use the strong law of large numbers to prove this statement. However, our approach strengthens the point that Sanov's theorem has consequences beyond its useful operational interpretation in hypothesis testing. The setting of hypothesis testing is in fact not even necessary to prove and describe Sanov's theorem. However, approaching this theory with a hypothesis testing framework is informative in the classical setting. And in the quantum case, some might say the analogous role of quantum hypothesis testing in a quantum Sanov theorem is even necessary [16, 31, 23].

II.2.1 Hypothesis testing

In this section we will introduce the basic formalism of hypothesis testing in the classical and finite dimensional setting. Among other asymptotic results in hypothesis testing, we state Stein's lemma and what we call the operational version of Sanov's theorem. The proof of Stein's lemma will be done later in Section II.2.3 where we will employ Sanov's theorem.

Suppose A is finite non-empty set and $|A| = d$. Let $p, q \in \text{Prob}(A)$ be two distinct probability distributions where we consider A as a measure space paired with the σ-algebra of all subsets of A. Then suppose there is a random variable X which takes values on A that is known to be distributed according to either p or q. That is to say the probability $X = a$ is either $p(a)$ for all $a \in A$ or $q(a)$ for all $a \in A$. In this context there are two *hypotheses* for the underlying distribution of the random variable X. The first hypothesis is that X is distributed according to p, and the second hypothesis is that X is distributed according to q. For historical reasons, two such hypotheses are colloquially called the null hypothesis H_0 and alternative hypothesis H_1, which is commonly written as

$$\begin{aligned} H_0 &: X \text{ is a random variable with law } p, \\ H_1 &: X \text{ is a random variable with law } q. \end{aligned} \tag{2.13}$$

In this setting there are two natural questions. Firstly, how can one conclude the underlying distribution of X? And secondly, what is the optimal way to determine if the null or alternative hypothesis is correct? The answer to the first question is motivated by Proposition II.2.1. Suppose that X is distributed according to p, then an infinite sequence $(x_1, x_2, \cdots) \in A^{\mathbb{N}}$ of observations made from i.i.d copies of X almost surely satisfies $\lim_{n\to} L_n(x)(a) = p(a)$ for all $a \in A$. Or stated otherwise, since pointwise convergence on A is equivalent to convergence in the Euclidean metric on $\text{Prob}(A)$ we have,

$$p^{\mathbb{N}}\Big(x \in A^{\mathbb{N}}; \lim_{n\to\infty} L_n(x) = p\Big) = 1. \tag{2.14}$$

Therefore, a sequence of observations $x \in A^{\mathbb{N}}$ sampled from i.i.d copies of X almost surely satisfies $\lim_{n\to} L_n(x) = p$ or $\lim_{n\to} L_n(x) = q$. Of course, this mathematical idealization of taking an infinite sequence of observations is physically impossible. In a realistic setting,

one is limited to a finite amount of observations, say of length $n \in \mathbb{N}$, and the challenge is then to deduce if a string of observations $x = (x_1, \cdots, x_n)$ was produced from a random variable with either law p or q. From our view there are two very natural approaches to do this. First, somewhat inspired by (2.14), is to accept the null hypothesis or alternative hypothesis based on if the empirical measure $L_n(x)$ is closer to p or q with respect to the Euclidean metric. Namely,

$$\begin{aligned} &\text{Accept } H_0: \text{ if } \sum_{a\in A} \big|L_n(x)(a) - p(a)\big|^2 \leq \sum_{a\in A} \big|L_n(x)(a) - q(a)\big|^2, \\ &\text{Accept } H_1: \text{ otherwise.} \end{aligned} \tag{2.15}$$

We can equivalently say, if $x \in E_n$ then we accept H_0, and if $x \in A^n \backslash E_n$ then we accept H_1, where

$$E_n := \Big\{ x \in A^n; \sum_{a\in A} \big|L_n(x)(a) - p(a)\big|^2 \leq \sum_{a\in A} \big|L_n(x)(a) - q(a)\big|^2 \Big\}. \tag{2.16}$$

The second natural approach we can devise is to accept H_0 if $x \in T_n^*$ and accept H_1 if $x \notin T_n^*$ where,

$$T_n^* := \{x \in A^n; p^n(x) \geq q^n(x)\}. \tag{2.17}$$

In other words, T_n^* is designed such that one accepts the hypothesis which was most likely to produce the sampled data $x = (x_1, \cdots, x_n)$. The two subsets $E_n \subseteq A^n$ and $T_n^* \subseteq A^n$ we have constructed are examples of a more general notion called *tests*.

Definition II.2.1 (Classical Tests)**.** *A subset $T_n \subset A^n$ is called a test if the basis for accepting the null or alternative hypothesis is whether $x \in T_n$ or $x \in A^n \backslash T_n$. That is to say, for some string of observations $x \in A^n$ a tests T_n specifies:*

$$\begin{aligned} &\text{if } x \in T_n \implies \text{accept } H_0, \\ &\text{if } x \notin T_n \implies \text{Accept } H_1. \end{aligned} \tag{2.18}$$

It turns out that the test T_n^* is a far more sensible test than E_n, which leads us to an answer to our second question. Recall, we wanted to know what is the most optimal way to determine if X is distrusted according to p or q? There are several approaches to defining what is an optimal test, and naturally, different standards of optimality require different tests. A central concept to nearly all discussions regarding the quality or optimality of a test $T_n \subseteq A^n$ is its error rates. In the hypothesis testing regime as detailed in (2.13), one classifies the errors for a test T_n as follows. Let $x = (x_1, \cdots, x_n) \in A^n$ be a string of n observations sampled from i.i.d distributed random variables $X_1, \cdots, X_n$ and $T_n \subseteq A^n$ be an arbitrary test. We say a false positive error occurs, or a type I error, if $x \notin T_n$ while X is distrusted according to p. Or stated otherwise, H_0 is in fact true while the underlying test T_n and string of observations x would make one conclude H_1 is the valid hypothesis. The probability of a type I error occurring is denoted by $\alpha(T_n)$ and given by

$$\alpha(T_n) := p^n\Big(A^n \backslash T_n\Big). \tag{2.19}$$

Similarly, we say a false negative occurs, or type II error, if $x \in T_n$ while X is distributed according to q. Or, the test T_n and data x would make one falsely conclude H_0 while H_1 is true. The probability of a type II error occurring is denoted by $\beta(T_n)$ and given by

$$\beta(T_n) := q^n\Big(T_n\Big). \tag{2.20}$$

In a real world setting there are two equally sensible ways to define if a test is optimal. The first approach, which we call optimality in terms of Chernoff, is to say a test $T_n \subset A^n$ is optimal if it minimizes the sum of the type I and type II errors, namely,

$$\gamma(T_n) := \alpha(T_n) + \beta(T_n). \tag{2.21}$$

Chernoff's optimality treats the false positive and false negative errors *symmetrically*.

Intermezzo
The definition of Chernoff's optimality is in fact better motivated by adding a slight bit of complexity to our hypothesis testing setup. We initially assumed that the random variable X either had law p, the null hypothesis, or law q, the alternative hypothesis. However, we can also specify the null hypothesis is correct with probability $t \in [0, 1]$ and the alternative hypothesis is correct with probability $1 - t$. While this generalization may seem a bit convoluted at first, it in fact clarifies the applicability of hypothesis testing. Suppose we have a finite family of random variables $\{X, Y, \cdots, Z\}$, all of which are either distributed according to p or q. Then by assumption a fraction $t \in [0, 1]$ of the random variables in $\{X, Y, \cdots, Z\}$ have law p and a fraction $1 - t$ have law q. Consider the following chain of events. Suppose one *randomly* selects a random variable X_0 in the set $\{X, Y, \cdots, Z\}$, where each random variable has an equal probability of being selected, then upon sampling n copies of X_0, the probability that the string of observations $x = (x_1, \cdots, x_n) \in A^n$ and test T_n will result in either a type I or type II error is given by

$$\gamma_t(T_n) := t \cdot \alpha(T_n) + (1 - t) \cdot \beta(T_n).$$

Namely, the probability one makes any type of error is the probability the selected random variable X has law p AND one incorrectly concludes X has law q, plus the probability that X has law q AND one incorrectly concludes X has law p. If $t = \frac{1}{2}$ then X_0 is equally likely to have law p or q, and the probability a given test $T_n \subseteq A^n$ produces any type of error is proportional to the Chernoff optimality of the test T_n,

$$2 \cdot \gamma_{\frac{1}{2}}(T_n) = \gamma(T_n).$$

And clearly any test that minimizes the right hand side is also a minimizer of the left hand side.
End of Intermezzo

Another approach, which we call optimality in terms of Stein, is to say a test $T_n \subset A^n$ is Stein optimal if it minimizes the type II error while ensuring the type I error is bounded above by some fixed $\delta > 0$. This notion of optimally is *asymmetric* in the sense that the event in which a type II error occurs has drastically greater consequences than the event in which a type I error occurs. Thus, minimizing the probability that a type II error occurs is more reasonable than minimizing both errors equally, as was done in Chernoff's case.

The vast applicability of hypothesis testing in science is already enough merit to learn about this material. However, the particular application we are most excited about is the operational interpretation that hypothesis testing provides of entropies in mathematical physics. Let $\delta > 0$ and consider the hypotheses described in (2.35) for a random variable X. Let $n \in \mathbb{N}$ and define

$$\beta^*_{n,\delta} := \inf\{\beta(T); T \subseteq A^n \text{ given } \alpha(T) < \delta\}, \tag{2.22}$$

which is the type II error for the Stein optimal test. On the other hand, define

$$\gamma_n{}^* := \inf\{\alpha(T) + \beta(T); T \subseteq A^n\}, \tag{2.23}$$

which is the total error for the Chernoff optimal test. Therefore, we have a satisfactory answer for what is the optimal way to deduce if the null or alternative hypothesis is correct. Namely, depending on the context a test $S^*_N \subseteq A^n$ can be selected that satisfies Stein's optimality or a test $T^*_n \subseteq A^n$ can be selected that satisfies Chernoff's optimality. For the case of Chernoff's optimality, we know a test T^*_n exists that achieves the infimum, $\gamma^*_n = \gamma(T^*_n)$, since the collection of subsets of A^n is finite. In fact, recall we defined the test T^*_n in equation (2.17), which turns out to be the test that achieves Chernoff's optimality.

Lemma II.2.1. *The test* $T^*_n = \{x \in A^n; p^n(x) \geq q^n(x)\}$ *achieves the minimum of* (2.23)*,*

$$\gamma(T^*_n) = \gamma_n{}^*.$$

The proof of this statement is simple and will be done in a more general setting in Lemma II.2.6. As for the existence of the Stein optimal test, we know for all $\delta > 0$ there exists at least one test S_N which satisfies $\alpha(S_N) < \delta$, namely, take $S_N = A^n$ then $\alpha(S_N) = p^n(\emptyset) = 0$. Hence, the infimum in (2.22) is taken over a finite non-empty set and for all $n \in \mathbb{N}$ there is a test S^*_N such that the type I error is bounded $\alpha(S^*_N) < \delta$ and, $\beta(S^*_N) = \beta^*_{n,\delta}$. It is important to remember the optimal errors rates as specified by Stein and Chernoff are only dependent on the null and alternative hypothesis, or more concretely,

$$\beta^*_{n,\delta} = \beta^*_{n,\delta}(p, q) \text{ and } \gamma^*_n = \gamma^*_n(p, q), \tag{2.24}$$

are both functions on $\text{Prob}(A) \times \text{Prob}(A)$. For two distinct probability distributions $p, q \in \text{Prob}(A)$. a topic of central importance is if we can capture the asymptotic behaviour of $\beta^*_{n,\delta}(p, q)$ and $\gamma^*_n(p, q)$ as the amount of observations increases $n \to \infty$. As expected, having more data reduces the probability that one can make either a false positive or a false negative error. The decay of the two error rates as defined for Stein's and Chernoffs optimal tests are quite beautifully captured by two different entropies.

Firstly, the Kullback-Leibler relative entropy describes the asymptotic decay of $\beta^*_{n,\delta}(p, q)$.

Lemma II.2.2 (Stein's Lemma)**.** *Let* $\varepsilon > 0$ *and* $p, q \in$ *Prob*(A)*, and define* $\beta^*_{n,\varepsilon}$ *as in* (2.22)*. Then*

$$\lim_{n\to\infty} \frac{1}{n} \log \beta^*_{n,\varepsilon} = -D_{KL}(p||q). \tag{2.25}$$

As previously mentioned we leave the proof of Stein's lemma until Section II.2.3. In a similar fashion the rate of decay of $\gamma^*_n(p, q)$ is captured by what we call Chernoff's entropy $C(p, q) : \text{Prob}(A) \times \text{Prob}(A) \to [0, \infty)$, which is defined as

$$C(p,q) := \sup_{\alpha\in(0,1)} (1-\alpha)R_\alpha(p||q), \tag{2.26}$$

where $R_\alpha(p||q)$ is the Rényi relative entropy given in (2.8).

Lemma II.2.3 (Chernoff's Lemma)**.** *Let $p, q \in Prob(A)$ and define γ_n^* as in* (2.23)*. Then*

$$\lim_{n\to\infty} \frac{1}{n} \log \gamma_n^* = -C(p,q). \tag{2.27}$$

Both Stein's and Chernoff's lemma give very similar operational interpretations of their respective entropies. Firstly, suppose $S_N^* \subset A^n$ is the sequence of tests which achieves Stein's optimality for a fixed $\delta > 0$; then for large n the probability that a type II error occurs for the sequence of tests S_N^* is estimated by

$$\exp(-n(D_{KL}(p||q) + \varepsilon)) \leq \beta(S_N^*) \leq \exp(-n(D_{KL}(p||q) - \varepsilon)), \tag{2.28}$$

where $\varepsilon \in (0,1)$ can be taken arbitrarily small, while obviously $N = N(\varepsilon)$ is a function of ε. Equivalently, if $T_n^* \subseteq A^n$ is the sequence of tests that achieves Chernoff's optimality $\gamma_n^* = \gamma(T_n^*)$, then

$$\exp(-n(C(p,q) + \varepsilon)) \leq \gamma(T_n^*) \leq \exp(-n(C(p,q) - \varepsilon)). \tag{2.29}$$

From both of these relations, we see that the respective relative entropies indicate how distinct two probability measures are. In the case of the Kullback Leibler relative entropy this interpretation can be taken further. Let $p, q, r \in \text{Prob}(A)$ be three different probability distributions and suppose X is a random variable with law q, while $x = (x_1, \cdots, x_n)$ is a string of observations producing by sampling n copies of X. If

$$D_{KL}(p||q) < D_{KL}(r||q), \tag{2.30}$$

the probability one mistakenly concludes the x was produced by a random variable with law p is greater than the probability one mistakenly concludes that x was produced by a random variable with law r. So q is more easily mistaken for p than r, and the Kullback-Leibler relative entropy encodes that behavior. It is in fact quite common for the Kullback-Leibler relative entropy to be referred to as the distance between two probability measures. While the Kullback-Leibler relative entropy does not satisfy the definition of a metric, namely it is not symmetric, it does successfully replace the role of the Euclidean metric in our first naive definition of the test E_n in (2.16). Recall, we wanted use the distance between the empirical measure $L_n(x)$ from p and q as means to determine if x was produced by a random variable with law p or q. By replacing the Euclidean metric with the Kullback-Leibler relative entropy, we can instead define the test

$$R_n := \{x \in A^n; D_{KL}(L_n(x)||p) \leq D_{KL}(L_n(x)||q)\}. \tag{2.31}$$

And in a rather unexpected fashion, the Kullback-Leibler relative entropy makes an appearance in Chernoff's lemma. Namely, let $x = (x_1, \cdots, x_N) \in A^n$ be a string of observations and suppose n_a is the number of time $a \in A$ appears in the string x, then

$$D_{KL}(L_n(x)||p) \leq D_{KL}(L_n(x)||q) \iff \sum_{a\in A} \frac{n_a}{n} \log p(a) \geq \sum_{a\in A} \frac{n_a}{n} \log q(a), \tag{2.32}$$

$$p^n(x) = \prod_{a \in A} p(a)^{n_a} \geq \prod_{a \in A} q(a)^{n_a} = q^n(x), \tag{2.33}$$

therefore we have

$$T_n^* = \{x \in A^n; p^n(x) \geq q^n(x)\} = R_n. \tag{2.34}$$

In Chernoff's lemma the role of a relative entropy is twofold. Firstly, under the assumption that it is equally important to not mistake p for q, or q for p, the Chernoff relative entropy captures the most effective way one can use observational data to distinguish p and q. Secondly, the Kullback-Leibler relative entropy can be used to define the Chernoff optimal test as shown in (2.31) and (2.34).

We finish of our section on classical hypothesis testing by stating what we call the operational Sanov theorem, which is essentially a generalization of of Stein's lemma. To state this result we will generalize our initial hypothesis testing setup. Recall we originally wanted to conclude if a random variable X is distributed according to p or q. This type of hypotheses are called atomic, since for both the null and alternative hypothesis, X can only be distributed according to one measure. Now let $B, C \subseteq \text{Prob}(A)$ be two disjoint sets of probability distributions and assume X has a distribution $\mu \in \text{Prob}(A)$ such that either $\mu \in B$ or $\mu \in C$. We can make the hypotheses,

$$\begin{aligned} &H_0 : X \text{ is distributed according to a measure in } B, \\ &H_1 : X \text{ is distributed according to a measure in } C. \end{aligned} \tag{2.35}$$

This type of hypothesis testing is known as compound hypothesis testing, and we say both the null and alternative hypothesis are compound. For a given test $T_n \subseteq A^n$ we again specify:

$$\begin{aligned} &\text{if } x \in T_n \implies \text{accept } H_0, \\ &\text{if } x \notin T_n \implies \text{Accept } H_1. \end{aligned} \tag{2.36}$$

We can extend the definition of a type I and type II error for a test $T_n \subseteq A^n$,

$$\alpha(T_n) := \sup\{p^n(A^n \backslash T_n); p \in B\}, \quad \beta(T_n) := \sup\{q^n(T_n); q \in C\}, \tag{2.37}$$

where we take the supremum over B and C to assess the test T_n under the worst possible circumstances. Note if $B = \{p\}$ and $C = \{q\}$, as is the case in simple hypothesis testing, then the previous definition of a type I and II error coincide with the two above. Moving on, for $\delta > 0$ we define the analogous Stein optimal error rate as

$$\beta_{n,\delta}^*(B, C) := \inf\{\beta(T_n); T_n \subseteq A^n \text{ and } \alpha(T_n) < \delta\}. \tag{2.38}$$

We can now state the extension of Stein's lemma to compound hypothesis testing.

Theorem II.2.1 (Operational Sanov Theorem)**.** *Fix $\delta > 0$ and let $B, C \subseteq Prob(A)$. For all $n \in \mathbb{N}$ define $\beta_{n,\delta}^*(B, C)$ as in* (2.38)*. Then*

$$\lim_{n \to \infty} \frac{1}{n} \log \beta_{n,\delta}^*(B, C) = - \inf_{p \in B, q \in C} D_{KL}(p||q). \tag{2.39}$$

We will postpone the sketch proof of Theorem II.2.1 to Section II.2.3. In the setting that $C = \{q\}$ is a singleton, the quantum extension of the operational Sanov theorem is often referred to as the quantum Sanov theorem. We will state and prove the quantum extension of Theorem II.2.1, however, we restrain from calling it a proper extension of the Sanov theorem, and rather consider it an extension of the operational Sanov theorem. At a later stage in this thesis we will prove a result in the quantum setting which we believe more closely resembles the classical Sanov theorem. However, first we must understand and prove the classical result.

II.2.2 Sanov's theorem

Suppose A is a finite non-empty set, in this section we focus on proving Sanov's theorem in the simplified setting where A, paired with σ-algebra of all subsets of A, is our underlying measure space. At a later stage we treat the case where the underlying space is instead Polish. In part, we first prove Sanov's theorem in the finite dimensional setting since we do not assume the reader is familiar with large deviation theory. Additionally, all efforts taken to extend the Sanov theorem to a quantum regime have also been limited to the finite dimensional setting. When making comparisons between the classical and quantum versions of the Sanov theorem, we believe one should limit the scope to two finite systems. We build on the proofs presented in [26] for most of our results. Fix an $n \in \mathbb{N}$ and equip $\mathrm{Prob}(A)$ with the weak topology, which in this case is equivalent to the Euclidean topology as described in (2.6). We will extensively be studying the empirical measure $L_n : A^n \to \mathrm{Prob}(A)$, which maps string of n observations to probability measures $L_n(x) \in \mathrm{Prob}_n(A) \subseteq \mathrm{Prob}(A)$, which is the subset defined as

$$\mathrm{Prob}_n(A) := \Big\{ p \in \mathrm{Prob}(A); \forall a \in A, \exists n_a \in \mathbb{N}, \text{ such that } p(a) = \frac{n_a}{n} \Big\}. \tag{2.40}$$

Note that $\mathrm{Prob}_n(A)$ can equivalently be defined as the image of the empirical measure L_n,

$$\mathrm{Prob}_n(A) = \{ L_n(x); x \in A^n \} = L_n(A^n). \tag{2.41}$$

Sanov's theorem captures the asymptotic behaviour of the empirical measure L_n when the number of observations $n \to \infty$. The behaviour of L_n is always described with respect to a fixed measure $q \in \mathrm{Prob}(A)$, in fact by using the empirical measure L_n and q we can induce a probability measure on the set $\mathrm{Prob}_n(A) = L_n(A^n)$, where for each $L_n(x) \in \mathrm{Prob}_n(A)$ one assigns the probability $p^n(x)$. Quite satisfyingly, the induced measure on the set $\mathrm{Prob}_n(A)$ can be extended to a probability measure over the set of all probability distribution $\mathrm{Prob}(A)$. That is, for any weakly measurable set $\Omega \subseteq \mathrm{Prob}(A)$ the measure of Ω is given by

$$q^n(L_n^{-1}(\Omega)) = q^n(L_n \in \Omega). \tag{2.42}$$

An important conceptual remark about $\mathrm{Prob}_n(A)$ is that the union of all $\mathrm{Prob}_n(A)$ is dense in $\mathrm{Prob}(A)$,

$$\Big(\bigcup_{n \in \mathbb{N}} \mathrm{Prob}_n(A) \Big)^- = \mathrm{Prob}(A), \tag{2.43}$$

We can prove this result with two strategies: one is an elementary approach using that $\mathbb{Q}$ is dense in $\mathbb{R}$, which we do in the appendix (Lemma A.4.2), and the other is an application of Proposition II.2.1. As for the latter argument, for all $p \in \text{Prob}(A)$ we know $\lim_{n\to\infty} L_n(x) = p$, for $p^{\mathbb{N}}$ almost every $x \in A^{\mathbb{N}}$. Now pick one $x \in A^n$ with the property that $\lim_{n\to\infty} L_n(x) = p$, then $L_n(x) \in \text{Prob}_n(A)$ for all $n \in \mathbb{N}$, thus p is in the closure of $\bigcup_{n\in\mathbb{N}} \text{Prob}_n(A)$. Keeping this in mind, for any weakly open $\Omega \subseteq \text{Prob}(A)$ we can conclude for large enough n,

$$L_n^{-1}(\Omega) \neq \emptyset. \tag{2.44}$$

A useful interpretation of $L_n^{-1}(\Omega) \subseteq A^n$ is that it is the collection of the string of observations that will most likely be produced from sampling any $q \in \Omega$ exactly n times. The important idea to keep in mind is that for all $n \in \mathbb{N}$ the Bernoulli measure q^n on the set A^n induces a distribution on $\text{Prob}(A)$ through the empirical measure L_n, and this framework is exceptionally useful. It is at this point that we can make our first asymptotic statement about the behaviour of the empirical measure.

Proposition II.2.2. *Let $\Omega \subseteq \text{Prob}(A)$ be a weakly open set and $q \in \Omega$ then,*

$$\lim_{n\to\infty} q^n(L_n \in \Omega) = 1. \tag{2.45}$$

Furthermore, there exists an $D > 0$ such that for all $0 < d < D$ and large enough n,

$$q^n(L_n \in \Omega) \geq 1 - \exp(-nd). \tag{2.46}$$

It should at this point come as no surprise Proposition II.2.2 is conveniently another consequence of the Sanov theorem. We therefore leave the proof of Proposition II.2.2 until the next section, and for now focus on its interpretation. In Proposition II.2.2 we first establish that for all $q \in \Omega$, the sets $L_n^{-1}(\Omega) \subseteq A^n$ are *typical*, precisely meaning that (2.45) holds. Moreover, if one samples a string $x = (x_1, \cdots, x_n)$ from a random variable with law $q \in \Omega$, then the probability that $L_n(x) \in \Omega$ as $n \to \infty$ will approach 1 exponentially fast. As of now, this is our only knowledge about the asymptotic behaviour of L_n with respect to sets $\Omega \subseteq \text{Prob}(A)$ that are weakly open and probability measure $q \in \Omega$. Keeping in mind we are interested in understanding the general asymptotic properties of the empirical measure, thus we can loosen the two assumptions we made in Proposition II.2.2. Namely, instead assume $\Omega \subseteq \text{Prob}(A)$ is weakly measurable, and take any probability measure $q \in \text{Prob}(A)$. At last, we are now at a point where we can state Sanov's theorem.

Theorem II.2.2 (Sanov's theorem)**.** *Suppose A is a finite set and $|A| \geq 2$. Let $q \in \mathrm{Prob}(A)$ and let $\Omega \subset \mathrm{Prob}(A)$ be weakly measurable. Then*

$$\begin{aligned} -\inf_{p\in\Omega^\circ} D_{KL}(p||q) &\leq \liminf_{n\to\infty} \frac{1}{n}\log q^n(L_n \in \Omega) \\ &\leq \limsup_{n\to\infty} \frac{1}{n}\log q^n(L_n \in \Omega) \leq -\inf_{p\in\Omega^-} D_{KL}(p||q), \end{aligned} \tag{2.47}$$

where Ω° is the interior of Ω and Ω^- the closure of Ω.

For now we will focus on proving Sanov's theorem, and leave most of the discussion about its consequences to Section II.2.3. We will make extensive use of Lemma II.2.4 and Lemma II.2.5, which we will soon state and prove. For both lemmas we will make use of the Shannon entropy $S : \mathrm{Prob}(A) \to [0, \infty)$.

Definition II.2.2 (Shannon Entropy)**.** *Let A be a finite and non-empty set. We define the Shannon entropy for $q \in \mathrm{Prob}(A)$ as*

$$S(q) := -\sum_{a\in A} q(a)\log q(a). \tag{2.48}$$

Lemma II.2.4. *Let $n \in \mathbb{N}$ and $p, q \in \mathrm{Prob}_n(A)$ then,*

$$q^n(L_n^{-1}(p)) \leq q^n(L_n^{-1}(q)). \tag{2.49}$$

Moreover, for every $q \in \mathrm{Prob}(A)$ and $p \in \mathrm{Prob}_n(A)$ if $p \ll q$ then

$$q^n(L_n = p) = \big|L_n^{-1}(p)\big| \exp\Big(- n(D_{KL}(p||q) + S(p))\Big). \tag{2.50}$$

Proof. Fix a $n \in \mathbb{N}$ and let $p, q \in \mathrm{Prob}_n(A)$ while $(n_a)_{a\in A} \in \mathbb{N}^d$ and $(m_a)_{a\in A} \in \mathbb{N}^d$ are the two unique vectors such that for all $a \in A$,

$$q(a) = \frac{n_a}{n} \qquad \text{and} \qquad p(a) = \frac{m_a}{n}. \tag{2.51}$$

We will use that $x \in L_n(p)$ iff each element $a \in A$ appears exactly m_a times in the string x. Keeping this in mind, the cardinality of $L_n(p)$ should in fact be equal to the number of ways n items can be distributed in $d = |A|$ different sets which sizes $m_1, \cdots, m_d$ where each $m_1, \cdots, m_d \in \{m_a; a \in A\}$ is unique. That number is the multinomial coefficient

$$\big|L_n^{-1}(p)\big| = n!\big(\prod_{a\in A} m_a!\big)^{-1}. \tag{2.52}$$

Using this relation we then obtain

$$\begin{aligned} q^n(L_n^{-1}(p)) &= \sum_{x\in L_n^{-1}(p)} q^n(x) = \sum_{x\in L_n^{-1}(p)} \prod_{a\in A} q(a)^{m_a} \\ &= \big|L_n^{-1}(p)\big| \prod_{a\in A} q(a)^{m_a} \\ &= n!\big(\prod_{a\in A} m_a!\big)^{-1} \prod_{a\in A} q(a)^{m_a}. \end{aligned} \tag{2.53}$$

In a near identical fashion we have

$$q^n(L_n^{-1}(q)) = n!\big(\prod_{a\in A} n_a!\big)^{-1}\prod_{a\in A} q(a)^{n_a}. \tag{2.54}$$

Therefore,

$$\begin{aligned}\frac{q^n(L_n^{-1}(p))}{q^n(L_n^{-1}(q))} &= \prod_{a\in A}\frac{n_a!}{m_a!}\Big(\frac{n_a}{n}\Big)^{m_a-n_a} = \prod_{a\in A}\frac{n_a!}{m_a!}(n_a)^{m_a-n_a}N^{n_a-m_a}\\ &= \prod_{a\in A}\frac{n_a!}{m_a!}(n_a)^{m_a-n_a}. \end{aligned}\tag{2.55}$$

Using that for all $n, m \in \mathbb{N}$,

$$\frac{n!}{m!}\cdot n^{m-n} \leq 1, \tag{2.56}$$

as proved in the appendix in Lemma A.4.1, we can conclude

$$q^n(L_n^{-1}(p)) \leq q^n(L_n^{-1}(q)). \tag{2.57}$$

As for the second statement, assume $q \in \mathrm{Prob}(A)$ and $p \in \mathrm{Prob}_n(A)$, and as before let $p(a) = \frac{m_a}{n}$ for all $a \in A$. Then,

$$\begin{aligned} q^n(L_n^{-1}(p)) &= \sum_{x\in L_n^{-1}(p)} q^n(x) = \sum_{x\in L_n^{-1}(p)}\prod_{a\in A} q(a)^{m_a}\\ &= \big|L_n^{-1}(p)\big|\prod_{a\in A}\exp\log(q(a)^{m_a}) = \big|L_n^{-1}(p)\big|\exp\sum_{a\in A} m_a\log q(a)\\ &= \big|L_n^{-1}(p)\big|\exp\Big(N\sum_{a\in A}p(a)\log q(a) - n\sum_{a\in A}p(a)\big(\log p(a) - \log q(a)\big)\Big)\\ &= \big|L_n^{-1}(p)\big|\exp\big(-n(D_{KL}(p||q) + S(p))\big). \end{aligned}\tag{2.58}$$

□

The next lemma we will show provides estimates for the cardinalities of $\mathrm{Prob}_n(A)$ and $L_n^{-1}(p)$ when $p \in \mathrm{Prob}_n(A)$.

Lemma II.2.5. *Let $n \in \mathbb{N}$ and $p \in \mathrm{Prob}_n(A)$ then,*

$$(1+n)^{-d}\exp(nS(p)) \leq \big|L_n^{-1}(p)\big| \leq \exp(nS(p))), \tag{2.59}$$

and

$$\Big|\mathrm{Prob}_n(A)\Big| \leq (1+n)^d, \tag{2.60}$$

where $|A| = d$.

Proof. We begin with the easiest case. Notice that

$$\mathrm{Prob}_n(A) \simeq \Big\{(n_1,\cdots,n_d)\in\mathbb{N}^d; \sum_{k=1}^{d} n_i = N\Big\}, \tag{2.61}$$

and trivially for any $k \in \{1 \cdots, N\}$, $n_k \in \{0, \cdots, n\}$. Thus there are at most $(N+1)$ possible values for each n_k and

$$\left|\text{Prob}_n(A)\right| \leq (n+1)^d.$$

Now we will show the estimates for the set $|L_n^{-1}(p)|$. We begin with the upper bound, let $x = (x_1, \cdots, x_n) \in L_n^{-1}(p)$, by assumption we know for all $a \in A$ there exists an $n_a \in \mathbb{N}$ such that $p(a) = \frac{n_a}{n}$. Since $L_n(x) = p$ the element $a \in A$ appears exactly n_a times in the string x. Thus,

$$\begin{aligned} p^n(x) &= \prod_{k=1}^{n} p(x_i) = \prod_{a \in A} p(a)^{n_a} \\ &= \prod_{a \in A} \exp\log\left(p(a)^{n_a}\right) = \exp\left(n \sum_{a \in A} \frac{n_a}{n} \log p(a)\right) \\ &= \exp(-nS(p)). \end{aligned} \tag{2.62}$$

implying

$$p^n(L_n^{-1}(p)) = \left|L_n^{-1}(p)\right| \exp(-nS(p)) \leq 1 \implies \left|L_n^{-1}(p)\right| \leq \exp(nS(p)). \tag{2.63}$$

For the upper bound,

$$\begin{aligned} 1 &= \sum_{x \in A^n} p^n(x) = \sum_{\mu \in \text{Prob}_n(A)} p^n(L_n^{-1}(\mu)) \\ &\leq \sum_{\mu \in \text{Prob}_n(A)} p^n(L_n^{-1}(p)) \\ &\leq (1+n)^d p^n(L_n^{-1}(p)) \\ &= (1+n)^d |L_n^{-1}(p)| \exp(-nS(p)), \end{aligned} \tag{2.64}$$

where in the first inequality we we apply (2.49), in the second inequality we use that $|\text{Prob}_n(A)| \leq (1+n)^d$, and in the last equality we use (2.62). □

Before we finalize the proof of Sanov's theorem, we will briefly discuss how this result is an example of a large deviation principle, as defined in the appendix (Definition A.2.3). Let $\mathcal{X} = \text{Prob}(A)$ and $\mathcal{G}$ be the σ-algebra of all weakly measurable subsets of $\mathcal{X}$. As discussed previously, for all $q \in \text{Prob}(A)$ and $n \in \mathbb{N}$ we can extend the Bernoulli measure q^n to a probability measure in $\text{Prob}(\mathcal{X})$ with the empirical measure L_n. If we denote $q^n \circ L_n^{-1} \in \text{Prob}(\mathcal{X})$ as the probability measure satisfying

$$q^n \circ L_n^{-1}(\Omega) := q^n(L_n \in \Omega), \tag{2.65}$$

for all $\Omega \in \mathcal{G}$, then Sanov's theorem equivalently states that the sequence of measure $(q^n \circ L_n^{-1})_{n \in \mathbb{N}}$ satisfies a **large deviation principle** on $\mathcal{X}$ with rate function

$$I(p) := D_{KL}(p||q), \text{ where } p \in \mathcal{X}. \tag{2.66}$$

Proof of Sanov's Theorem II.2.2.

We will prove that for all open $U \subseteq \text{Prob}(X)$ and closed $V \subseteq \text{Prob}(X)$,

$$-\inf_{p\in U} D_{KL}(p||q) \leq \liminf_{n\to\infty} \frac{1}{n} \log q^n(L_n), \tag{2.67}$$

$$-\inf_{p\in V} D_{KL}(p||q) \geq \limsup_{n\to\infty} \frac{1}{n} \log q^n(L_n). \tag{2.68}$$

By Proposition A.2.1 this condition is equivalent to (2.47) and thus proves Sanov's theorem. We begin by showing (2.67). Let $U \subseteq \text{Prob(A)}$ be open and assume

$$\inf_{p\in U} D_{KL}(p||q) < \infty. \tag{2.69}$$

If $\inf_{p\in U} D_{KL}(p||q) = \infty$ then the inequality trivially holds. Let $\varepsilon > 0$, by definition of the infimum we know there exists a $p \in U$ such that

$$D_{KL}(p||q) - \varepsilon < \inf_{p\in U} D_{KL}(p||q). \tag{2.70}$$

Furthermore, there exists a sequence of probability measures $(p_n)_{n\in\mathbb{N}}$ such that $p_n \in \text{Prob}_n(A)$, $p_n \ll p$ and $p_n \to p$ as $n \to \infty$, where we use

$$\Big(\bigcup_{n\in\mathbb{N}} \text{Prob}_n(A)\Big)^- = \text{Prob}(A), \tag{2.71}$$

as proven in Lemma A.4.2. In general, the Kullback-Leibler relative entropy with respect to q is continuous on the domain

$$D_{KL}(\cdot||q) : \{p \in \text{Prob}(A); p \ll q\} \to [0, \infty). \tag{2.72}$$

This immediately follows from the fact that the function

$$(x, y) \mapsto x \log x - x \log y, \tag{2.73}$$

is continuous on $(0, 1) \times (0, 1) \subseteq \mathbb{R}^2$. By assumption $p \in U$, where U is open, so there exists a $K \in \mathbb{N}$ such that for all $n \geq K$, $p_n \in U$. Moreover, if $a \in A$ and $q(a) = 0$, then we can assume for all $n \in \mathbb{N}$, $p_n(a) = 0$, namely, the sequence of probabilities are absolutely continuous with respect to p and therefore also q. Since the Kullback-Leibler relative entropy is continuous on the set $\{p \in \text{Prob}(A); p \ll q\}$, we can pick an $M \in \mathbb{N}$ such that for all $n \geq M$,

$$\big|D_{KL}(p_n||q) - D_{KL}(p||q)\big| < \varepsilon. \tag{2.74}$$

Let $n \geq \max\{M, K\}$ by (2.50) in Lemma II.2.4 we have

$$\begin{aligned}\frac{1}{n} \log q^n(L_n) &\geq \frac{1}{n} \log q^n(L_n = p_n) \\ &= \frac{1}{n} \log \Big\{ \big|L_n^{-1}(p_n)\big| \exp\Big(- n(D_{KL}(p_n||q) + S(p_n)\Big)\Big\}. \end{aligned} \tag{2.75}$$

Next we apply the fact that the function $x \mapsto \log x$ is increasing on $\mathbb{R}+$ along with (2.59) in Lemma II.2.5,

$$\begin{aligned}
\frac{1}{n}\log q^n(L_n) &\geq \frac{1}{n}\log\Big\{(1+n)^{-d}\exp(nS(p_n))\exp\Big(-n(D_{KL}(p_n||q)+S(p_n)\Big)\Big\} \\
&= \frac{1}{n}\log\Big\{(1+n)^{-d}\exp\big(-nD_{KL}(p_n||q)\big)\Big\} \\
&= -\frac{d}{n}\log(1+n) - D_{KL}(p_n||q) \\
&\geq -\frac{d}{n}\log(1+n) - D_{KL}(p||q) - \varepsilon \\
&\geq -\frac{d}{n}\log(1+n) - 2\varepsilon - \inf_{p\in U} D_{KL}(p||q). \qquad (2.76)
\end{aligned}$$

Upon taking the limit inferior of both sides and then letting $\varepsilon \downarrow 0$ we obtain (2.67). Now we will prove (2.68), which turns out to be much simpler. Let $V \subseteq \mathrm{Prob}(A)$ be closed and $n \in \mathbb{N}$, then by (2.50),

$$\begin{aligned}
q^n(L_n) &= \sum_{p\in V\cap \mathrm{Prob}_n(A)} q^n(L_n = p) \\
&= \sum_{\substack{p\ll q;\\ p\in V\cap \mathrm{Prob}_n(A)}} \big|L_n^{-1}(p)\big|\exp\big(-n(D_{KL}(p||q)+S(p))\big), \qquad (2.77)
\end{aligned}$$

where we use the fact that $q^n(L_n^{-1}(p)) = 0$ if $p \not\ll q$, namely if there exists an $a \in A$ such that $q(a) = 0$ and $p(a) = \frac{m_a}{n} \neq 0$, then any string $x = (x_1, \cdots, x_n) \in L_n^{-1}(p)$ will necessarily contain m_a elements with the value a and $q^n(x) = 0$. By applying (2.59) we have

$$\begin{aligned}
q^n(L_n) &= \sum_{\substack{p\ll q;\\ p\in V\cap \mathrm{Prob}_n(A)}} \big|L_n^{-1}(p)\big|\exp\big(-n(D_{KL}(p||q)+S(p))\big) \\
&\leq \sum_{\substack{p\ll q;\\ p\in V\cap \mathrm{Prob}_n(A)}} \exp(nS(p))\exp\big(-n(D_{KL}(p||q)+S(p))\big) \\
&\leq \sum_{\substack{p\ll q;\\ p\in V\cap \mathrm{Prob}_n(A)}} \exp\big(-n\inf_{p\in V} D_{KL}(p||q)\big), \qquad (2.78)
\end{aligned}$$

where we use the fact that $x \mapsto \exp(-x)$ is a decreasing function on $\mathbb{R}^+$. Finally, we apply (2.60) in Lemma II.2.5 to obtain

$$\begin{aligned}
\frac{1}{n}\log q^n(L_n) &\leq \frac{1}{n}\log\Big\{(1+n)^d\exp\big(-n\inf_{p\in V} D_{KL}(p||q)\big)\Big\} \qquad (2.79) \\
&= \frac{d}{n}\log(1+n) - \inf_{p\in V} D_{KL}(p||q), \qquad (2.80)
\end{aligned}$$

and by taking the limit superior of both sides we get (2.68). □

II.2.3 Implications of Sanov's theorem

Throughout this chapter we have postponed a handful of proofs, including a proof of Stein's lemma and several asymptotic results regarding the behaviour of the empirical measure $L_n : A^n \to \text{Prob}(A)$. The main motivation for doing this was that we wanted to show that in principle all these results follow from Sanov's theorem. While it is not strictly necessary to use Sanov's theorem to prove any of the results, we like to explicitly showcase how the asymptotic behaviour of the empirical measure is a fundamental result in probability theory. And therefore, in the context of quantum probability it is very natural to investigate if there is any proper quantum extension of Sanov's theorem. The proceeding chapters in this thesis are all dedicated to this investigation, and a result from our research is that it is in fact not possible to have a completely analogous extension of Sanov's theorem to the quantum setting. To appreciate how this result establishes key differences between classical and quantum probability, we must first understand some consequences of Sanov's theorem in the classical setting. There are two technical statements we will make use of throughout this section:

Proposition II.2.3. *Let $q \in \text{Prob}(A)$ and $V \subseteq \text{Prob}(A)$ be a weakly closed set such that $q \notin V$. Then*

1. $$\inf_{p \in V} D_{KL}(p||q) > 0. \tag{2.81}$$

2. *Assume for all $a \in A$ that $q(a) > 0$, and suppose $U \subseteq \text{Prob}(A)$ is a weakly open or closed set. Then*

$$\inf_{p \in U^\circ} D_{KL}(p||q) = \inf_{p \in U^-} D_{KL}(p||q) < \infty. \tag{2.82}$$

Proof. 1. Suppose $\inf_{p \in V} D_{KL}(p||q) = 0$, then by definition of the infimum there exists a sequence $(p_n)_{n \in \mathbb{N}} \subseteq V$ such that

$$\lim_{n \to \infty} D_{KL}(p_n||q) = 0. \tag{2.83}$$

Then by a straight forward application of Pinsker's inequality (Proposition A.2.2) there exists an $C > 0$ such that,

$$||p_n - q|| \leq C\sqrt{D_{KL}(p_n||q)}, \quad \forall n \in \mathbb{N}. \tag{2.84}$$

Therefore $p_n \to q$ as $n \to \infty$, and since V is closed that would imply $q \in V$ which leads to our contradiction.

2. Since we assume for all $a \in A$ that $q(a) > 0$, the Kullback-Leibler relative entropy with respect to q

$$D_{KL}(\cdot||q) : \text{Prob}(A) \to [0, \infty), \tag{2.85}$$

is continuous on the whole set $\text{Prob}(A)$. Therefore $U^- \subseteq \text{Prob}(A)$ is a closed and bounded set and by the Heine-Borel theorem it is compact. Finally by the extreme value theorem,

$$\inf_{p \in U^\circ} D_{KL}(p||q) = \inf_{q \in U^-} D_{KL}(p||q). \tag{2.86}$$

That is, the infimum over U^- is achieved by some $p^* \in U^-$, and there is as sequence $(p_n)_{n\in\mathbb{N}} \subseteq U$ that converges to p^*. □

We return to proving Proposition II.2.1 which in fact is another formulation of the strong law of large numbers stated in Theorem A.2.4. Recall that a for a non-empty finite set A we defined $\mathcal{F}$ to be the σ-algebra generated by cylinder sets (2.10) in $A^{\mathbb{N}}$. Then, for some probability measure $q \in \text{Prob}(A)$, Proposition II.2.1 states that for $q^{\mathbb{N}}$ almost every $x \in A^{\mathbb{N}}$, we have $\lim_{n\to\infty} L_n(x) = q$, where the convergence is with respect to the weak topology, or equivalently pointwise convergence on A given $|A| < \infty$.

Proof of Proposition II.2.1. Let $q \in \text{Prob}(A)$ and define $A_0 := \{a \in A; q(a) > 0\}$, we can then consider $q \in \text{Prob}(A_0)$, and we will apply Sanov's theorem in this setting. For $\varepsilon > 0$ and $p \in \text{Prob}(A_0)$ define the two open sets

$$V_{0,\varepsilon}(p) := \{h \in \text{Prob}(A_0); ||h - p|| > \varepsilon\}, \tag{2.87}$$

$$B_{0,\varepsilon}(p) := \{h \in \text{Prob}(A_0); ||h - p|| < \varepsilon\}, \tag{2.88}$$

where by construction $B_{0,\varepsilon}(p)^c = V_{0,\varepsilon}(p)^-$. By Proposition II.2.3 we then know

$$\inf_{p\in V_{0,\varepsilon}(q)} D_{KL}(p||q) = \inf_{p\in V_{0,\varepsilon}(q)^-} D_{KL}(p||q) = \inf_{p\in B_{0,\varepsilon}(q)^c} D_{KL}(p||q) < \infty, \tag{2.89}$$

and Sanov's Theorem II.2.2 states

$$\lim_{n\to\infty} \frac{1}{n} \log q^n\big(L_n^{-1}\big(B_{0,\varepsilon}(q)^c\big)\big) = -\inf_{p\in B_{0,\varepsilon}(q)^c} D_{KL}(p||q). \tag{2.90}$$

Since $B_{0,\varepsilon}(q)^c$ is closed and $q \notin B_{0,\varepsilon}(q)^c$, by Proposition II.2.3

$$\inf_{p\in B_{0,\varepsilon}(q)^c} D_{KL}(p||q) = D > 0. \tag{2.91}$$

Therefore for $\delta > 0$ there exists an $M \in \mathbb{N}$ such that for all $n \geq M$,

$$1 - q^n(L_{n0,\varepsilon}(q)) = q^n(L_{n0,\varepsilon}(q)^c) \leq \exp\big(-n(D - \delta)\big). \tag{2.92}$$

Let $n \in \mathbb{N}$, recall the empirical measure $L_n : A_0^{\mathbb{N}} \to \text{Prob}(A_0)$ can be seen as a function on $A_0^{\mathbb{N}}$. Define

$$\begin{aligned} K_{0,n} :=& L_n^{-1}(B_{0,\varepsilon}(q)^c) = \{x \in A_0^{\mathbb{N}}; L_n(a) \in B_{0,\varepsilon}(q)^c\} \\ =& \bigcup_{a_1\in A_0} \cdots \bigcup_{a_{n0}} \Big\{[a_1, \cdots, a_n]; \exists x \in [a_1, \cdots, a_n] \text{ s.t } L_n(x) \in B_{0,\varepsilon}(q)^c\Big\}. \end{aligned} \tag{2.93}$$

The second equality holds above since for any set $U_0 \subseteq \text{Prob}(A_0)$, if $a_1, \cdots, a_{n0}$ and there exists an $b \in [a_1, \cdots, a_n]$ such that $L_n(b) \in U_0$, then for all $x \in [a_1, \cdots, a_n]$, $L_n(x) \in U_0$. Therefore, in probability space $(A_0^{\mathbb{N}}, \mathcal{F}_0, q^{\mathbb{N}})$ the sets $(K_{0,n})_{n\in\mathbb{N}} \subseteq \mathcal{F}_0$ are measurable and by (2.92),

$$\sum_{n=1}^{\infty} q^n(L_{n0,\varepsilon}(q)^c) = \sum_{n=1}^{\infty} q^{\mathbb{N}}(K_{0,n}) < \infty. \tag{2.94}$$

Finally, by Borel-Cantelli's Theorem A.2.3 we have

$$q^{\mathbb{N}}\Big(\Big\{a \in A_0^{\mathbb{N}}; L_n(a) \in B_{0,\varepsilon}(q)^c \text{ infinitely often}\Big\}\Big) = 0, \tag{2.95}$$

or stated otherwise, for $q^{\mathbb{N}}$ almost every $a \in A_0^{\mathbb{N}}$ we have $L_n(a) \notin B_{0,\varepsilon}(q)$ for at most finitely many n, so there exists an $M \in \mathbb{N}$ such that for all $n \geq M$, $L_n(a) \in B_{0,\varepsilon}(q)$. Because $\varepsilon > 0$ was taken to be arbitrary, we can conclude that for $q^{\mathbb{N}}$ almost every $x \in A_0^{\mathbb{N}}$,

$$\lim_{n\to\infty} L_n(x) = q, \tag{2.96}$$

where the convergence is with respect to the weak topology on $\text{Prob}(A_0)$. Now we will extend this result to the measure space $(A^{\mathbb{N}}, \mathcal{F}, q^{\mathbb{N}})$. Notice for all $k \in \mathbb{N}$ and $a_1, \cdots, a_k \in A_0 \subseteq A$ we have,

$$[a_0, \cdots, a_k] \subseteq \mathcal{F}, \tag{2.97}$$

which implies $\mathcal{F}_0 \subseteq \mathcal{F}$. Moreover, for all events $[a_1, \cdots, a_k] \in \mathcal{F}$,

$$q^n([a_1, \cdots, a_k]) > 0 \iff [a_1, \cdots, a_k] \in \mathcal{F}_0, \tag{2.98}$$

and for all $B \in \mathcal{F}$,

$$q^{\mathbb{N}}\big(B) = q^{\mathbb{N}}(B \cap A_0^{\mathbb{N}}) + q^{\mathbb{N}}(B \backslash A_0^{\mathbb{N}}) \geq q^{\mathbb{N}}(B \cap A_0^{\mathbb{N}}). \tag{2.99}$$

Now define the open ball

$$B_\varepsilon(q) := \{p \in \text{Prob}(A); ||p - q|| < \varepsilon\}, \tag{2.100}$$

then if we view $\text{Prob}(A_0) \subseteq \text{Prob}(A)$ as the set of probability distributions whose support is contained in A_0, we also have that

$$B_{0,\varepsilon}(q) \subseteq B_\varepsilon(q). \tag{2.101}$$

Hence,

$$\begin{aligned} 1 \geq & q^{\mathbb{N}}\Big(a \in A^{\mathbb{N}}; L_n(a) \notin B_\varepsilon(q) \text{ for at most finitely many } n\Big) \\ \geq & q^{\mathbb{N}}\Big(a \in A_0^{\mathbb{N}}; \exists M \in \mathbb{N} \text{ such that } \forall n \geq M, L_n(a) \in B_{0,\varepsilon}(q)\Big) = 1. \end{aligned} \tag{2.102}$$

Since $\varepsilon > 0$ was arbitrary we then get for $q^{\mathbb{N}}$ almost every $x \in A^{\mathbb{N}}$ that

$$\lim_{n\to\infty} L_n(x) = q. \tag{2.103}$$

□

It should be pointed out that we also largely proved Proposition II.2.2 in our argument above. The details are as follows.

Proof of Proposition II.2.2. . Let $q \in \text{Prob}(A)$ and $\Omega \subseteq \text{Prob}(A)$ an open set containing q. Let $A_0 = \{a \in A; q(a) > 0\}$, as in the previous proof we will consider $q \in \text{Prob}(A_0) \subseteq \text{Prob}(A)$ and apply Sanov's theorem in this setting. Because Ω is open there exists an $\varepsilon > 0$ such that $B_\varepsilon(q) \subseteq \Omega$, where $B_\varepsilon(q)$ is defined in (2.100). Furthermore, $q \in B_{0,\varepsilon}(q)$ which was the open ball in $\text{Prob}(A_0)$. Clearly $B_\varepsilon(q)^c$ is a closed set in $\text{Prob}(A)$ and $q \notin B_\varepsilon(q)^c$, similarly $B_{0,\varepsilon}(q)^c$ is closed in $\text{Prob}(A_0)$. By (2.92), for all $0 < d < D = \inf_{p \in B_{0,\varepsilon}(q)^c} D_{KL}(p||q)$ there exists an $M \in \mathbb{N}$ such that all all $n \geq M$,

$$q^n(L_{n0,\varepsilon}(q)) \geq 1 - \exp(-nd), \tag{2.104}$$

where crucially we view $q \in \text{Prob}(A_0)$, which is necessary since we are indirectly applying Proposition II.2.3. We can view $B_{0,\varepsilon}(q) = B_\varepsilon(q) \cap \text{Prob}(A_0) \subseteq \text{Prob}(A)$, and then we finish the proof with the inequality,

$$q^n(L_n \in \Omega) \geq q^n(L_{n\varepsilon}(p)) \geq q^n(L_{n0,\varepsilon}(q)). \tag{2.105}$$

□

So far we have used Sanov theorem to capture two similar results capturing the limiting behaviour of the empirical measure $L_n : A^{\mathbb{N}} \to \text{Prob}(A)$. Proposition II.2.1 puts us in a very neat setting. Namely, if $(x_n)_{n\in\mathbb{N}} \in A^{\mathbb{N}}$ is a sequence of observations made from an i.i.d sequence of random variables $(X_n)_{n\in\mathbb{N}}$ with law $q^{\mathbb{N}}$ then with probability equal to 1 the limit $\lim_{n\to\infty} L_n(x) = L_\infty(x)$ exists and $L_\infty(x) = q$. It is interesting to point out the limit $L_\infty(x)$ is generally not well defined. Even in the simple case $A = \{0,1\}^{\mathbb{N}}$ there are sequences $y \in A^{\mathbb{N}}$ such that

$$\liminf_{n\to\infty} L_n(y)(1) \neq \limsup_{n\to\infty} L_n(y)(1). \tag{2.106}$$

It is quite extraordinary that Kolmogorov's extension theorem along with Proposition II.2.1 captures (up to set of measure zero) the collection of all sequences $x \in A^{\mathbb{N}}$ for which the limit $L_\infty(x)$ is well defined. In the special context of studying i.i.d random variables, this observation sheds some light on how we should think about randomness in mathematics. That is, for a sequence $x \in A^{\mathbb{N}}$ if $\liminf_{n\to\infty} L_n(a) \neq \limsup_{n\to\infty} L_n(a)$ for some $a \in A$, then x cannot meaningfully be considered to be a random event produced by sampling an i.i.d sequence of random variables $(X_i)_{i\in\mathbb{N}}$ on A. The rather counter-intuitive part of this argument is for all $x \in A^{\mathbb{N}}$, the probability of sampling that sequence is zero, that is for all $p \in \text{Prob}(A)$, $p^{\mathbb{N}}(x) = 0$ even if $L_\infty(x) = p$. In [25] this anomaly is referred to as the *paradox of probability*, and we refer the reader to this work for a more detailed discussion about randomness in mathematics and physics.

To continue with our program, Proposition II.2.2 provides a more operational result than Proposition II.2.1. Namely, from Proposition II.2.2 for a finite string of observations $x = (x_1, \cdots, x_n)$ the probability that the empirical measure $L_n(x)$ is in some ε neighbourhood around q, converges exponentially fast to 1 as $n \to \infty$. This result has an immediate consequence in the hypothesis testing framework. Recall, given two distinct probability measures $p, q \in \text{Prob}(A)$ in (2.16) we defined the test

$$E_n := \big\{x \in A^n; ||L_n(x) - p|| < ||L_n(x) - q||\big\}. \tag{2.107}$$

We can reconstruct this test as follows. Define $\Omega_E \subseteq \text{Prob}(A)$ as

$$\Omega_E := \{h \in \text{Prob}(A); ||h - p|| < ||h - q||\}. \tag{2.108}$$

Then for all $n \in \mathbb{N}$,

$$E_n = L_n^{-1}(\Omega_E) \quad \text{and} \quad A^n \backslash E_n = L_n^{-1}(\Omega_E^c). \tag{2.109}$$

The type I and type II error probabilities are now given by

$$\alpha(E_n) = p^n\big(L_n \in \Omega_E^c\big) \quad \text{and} \quad \beta(E_n) = q^n\big(L_n \in \Omega_E\big). \tag{2.110}$$

Since the set Ω_E is open and $p \in \Omega_E$, by Proposition II.2.2 we know there is some $d > 0$ such that for large enough n,

$$p^n(L_n \in \Omega_E) = 1 - \alpha(E_n) \geq 1 - \exp(-nd), \tag{2.111}$$

or for all $\varepsilon > 0$ there exist an $M \in \mathbb{N}$ such that for all $n \geq M$,

$$\alpha(E_n) < \varepsilon. \tag{2.112}$$

Hence, for all $n \geq M$ it follows $\beta(E_n) \geq \beta^*_{\varepsilon,n}$ where $\beta^*_{\varepsilon,n}$ was the the Stein optimal rate defined in (2.22). And because $q \notin \Omega_E^-$,

$$\inf_{h \in \Omega_E} D_{KL}(h||q) \geq \inf_{h \in \Omega_E^-} D_{KL}(h||q) > 0. \tag{2.113}$$

If we set $d_1 = \inf_{h \in \Omega_E} D_{KL}(h||q)$ and $d_2 = \inf_{h \in \Omega_E^-} D_{KL}(h||q)$, then by Sanov's Theorem II.2.2,

$$-d_1 \leq \liminf_{n \to \infty} \frac{1}{n} \log \beta(E_n) \leq \limsup_{n \to \infty} \frac{1}{n} \log \beta_n(E_n) \leq -d_2. \tag{2.114}$$

In the case $q(a) > 0$ for all $a \in A$, then $d_1 = d_2$ and we have the more palatable result

$$\lim_{n \to \infty} \frac{1}{n} \log \beta_n(E_n) = - \inf_{h \in \Omega_E} D_{KL}(h||q). \tag{2.115}$$

It should not come as a surprise that $D_{KL}(p||q) > \inf_{h \in \Omega_E} D_{KL}(h||q)$, which follows from the strict convexity of the Kullback-Leibler relative entropy. Thus the test E_n is not a realization of the Stein optimal test S^*_n which would need to satisfy $\beta_n(S^*_n) = \beta^*_{n,\varepsilon}$, but it does confirm that the naive test E_n constructed from an intuitive perspective is by no means a poor choice. Moreover, the strategy of constructing a suitable subset $\Omega_T \subseteq \text{Prob}(A)$ such that

$$\beta(T_n) = q^n(L_n \in \Omega_T), \quad \text{and} \quad \alpha(T_n) = p^n(L_n \in \Omega_T^c), \tag{2.116}$$

for specific sequence of tests $T_n \subseteq A^n$ is the essential step we need to prove both Stein's Lemma II.2.2 and Chernoff's Lemma II.2.3 from Sanov's theorem. In the proof of these statements the following inequalities about the type I and II errors will be useful:

Lemma II.2.6. *Let $n \in \mathbb{N}$ and take any $\lambda \in \mathbb{R}$. Define the test $T^*_n(\lambda) \subseteq A^n$ as*

$$T^*_n(\lambda) := \Big\{ a \in A^n ; p^n(a) > \exp(n\lambda) q^n(a) \Big\}, \tag{2.117}$$

then for any test $Q_n \subseteq A^n$ we have

$$\alpha(Q_n) + e^{n \cdot \lambda} \beta(Q_n) \geq \alpha(T^*_n(\lambda)) + e^{n \cdot \lambda} \beta(T^*_n(\lambda)). \tag{2.118}$$

Proof. If $Q_n \subseteq A^n$, then

$$\begin{aligned}\alpha(Q_n) + e^{n\cdot\lambda}\beta(Q_n) &= p^n(Q_n^c) + e^{n\cdot\lambda}q^n(Q_n) = 1 - p^n(q^n) + e^{n\cdot\lambda}q^n(Q_n)\\ &= 1 + \sum_{\sigma\in Q_n}\Big(e^{n\cdot\lambda}q^n(\sigma) - p^n(\sigma)\Big)\\ &\geq 1 + \sum_{\sigma\in T_n^*(\lambda)}\Big(e^{n\cdot\lambda}q^n(\sigma) - p^n(\sigma)\Big)\\ &= \alpha(T_n^*) + e^{n\cdot\lambda}\beta(T_n^*) \end{aligned} \tag{2.119}$$

□

Proof of Stein's Lemma II.2.2. We will show Stein's lemma is a consequence Sanov's theorem in three steps. We will focus on proving Stein's lemma for $p, q \in \mathrm{Prob}(A)$ such that $p \ll q$ and p is not a Dirac distribution. The situations $p \not\ll q$ and $p = \delta_a$ for some $a \in A$ are much simpler and we omit proving these cases.

Step 1: Consider the hypothesis testing setup as described in (2.13). Let $\varepsilon > 0$ and $n \in \mathbb{N}$. We recall the definition of $\beta_{n,\varepsilon}^*$ in (2.22), which was given as

$$\beta_{n,\varepsilon}^* = \inf\{\beta(T); T \subseteq A^n \text{ given } \alpha(T) < \varepsilon\}. \tag{2.120}$$

Let $0 < \delta < \varepsilon$ be arbitrary, in this step we show the useful fact that for large enough n,

$$\beta_{n,\varepsilon-\delta}^* = \inf_{\lambda\in\mathbb{R}}\{\beta(T_n^*(\lambda)); \alpha(T_n^*(\lambda)) < \varepsilon\}, \tag{2.121}$$

meaning one can take the infimum over the specific tests $(T_n^*(\lambda))_{\lambda\in\mathbb{R}} \subseteq A^n$ as opposed to all tests. To prove this we will make use of Lemma II.2.6. Take an arbitrary test $Q_n \subseteq A^n$, such that $\alpha(Q_n) < \varepsilon - \delta$, then by Lemma II.2.6 we have for all $\lambda \in \mathbb{R}$,

$$\alpha(Q_n) + e^{n\cdot\lambda}\beta(Q_n) \geq \alpha(T_n^*(\lambda)) + e^{n\cdot\lambda}\beta(T_n^*(\lambda)). \tag{2.122}$$

Therefore, if there exists a $\lambda \in \mathbb{R}$ such that $\alpha(Q_n) \leq \alpha(T_n^*(\lambda)) < \varepsilon$, then necessarily $\beta(Q_n) \geq \beta(T_n^*(\lambda))$. Since the test Q_n is arbitrary (2.121) follows. To prove the existence of a $\lambda \in \mathbb{R}$ such that $\alpha(Q_n) \leq \alpha(T_n^*(\lambda)) < \varepsilon$, we will prove for any subinterval $(x_1, x_2) \subset (0, 1)$ there exists an $M \in \mathbb{N}$ such that for all $n \geq M$ there exists an $\lambda \in \mathbb{R}$ dependent on n, such that $\alpha(T_n^*(\lambda)) \in (x_1, x_2)$.
Notice for all $\lambda \in \mathbb{R}$,

$$\alpha(T_n^*(\lambda)) = p^n\Big(\Big\{a \in A^n; p^n(a) \leq \exp(n\lambda)q^n(a)\Big\}\Big) \tag{2.123}$$

$$\leq |A^n|e^{n\lambda}, \tag{2.124}$$

so as $\lambda \to -\infty$ we have $\alpha(T_n^*(\lambda)) \to 0$. On the other hand, for all $a \in A^n$, $p^n(a) \leq 1$, so clearly for $\lambda \to \infty$ we have $\alpha(T_n^*(\lambda)) \to 1$. Define the function $\Delta_n : \mathbb{R} \to [0, 1]$ as

$$\Delta_n(\lambda) := p^n\Big(\Big\{a \in A^n; p^n(a) = \exp(n\lambda)q^n(a)\Big\}\Big), \tag{2.125}$$

because the set A^n is finite and $p \ll q$; if $p^n(a) > 0$ then there exists a $\lambda_0 \in \mathbb{R}$ such that $p^n(a) = \exp(n\lambda_0)q^n(a)$. Hence if

$$L := \Big|\Big\{a \in A^n; p^n(a) \leq \exp(n\lambda)q^n(a)\Big\}\Big|, \tag{2.126}$$

there exists $\lambda_1, \cdots, \lambda_L \in \mathbb{R}$ where $\lambda_i \leq \lambda$ for $i \in \{1, \cdots, M\}$, such that

$$\alpha(T_n^*(\lambda)) = \sum_{i=1}^{L} \Delta_n(\lambda_i). \tag{2.127}$$

The key property to prove is that uniformly for all $\lambda \in \mathbb{R}$,

$$\lim_{N\to\infty} \Delta_n(\lambda) = 0, \tag{2.128}$$

that is, for any $\gamma > 0$ there exists an $M \in \mathbb{N}$ such that for all $n \geq M$ and all $\lambda \in \mathbb{R}$,

$$\Delta_n(\lambda) < \gamma. \tag{2.129}$$

Assume the uniform converge property is true, set $\gamma = \frac{x_2 - x_1}{2}$ and consider n large enough such that for all $\lambda \in \mathbb{R}$ (2.129) holds. Using that $\lim_{\lambda\to-\infty} \alpha(T_n^*(\lambda)) = 0$ we can pick a $\lambda_- \in \mathbb{R}$ and $\lambda_+ \in \mathbb{R}$ which are both independent of n such that $\alpha(T_n^*(\lambda_-)) = 0$ and $\alpha(T_n^*(\lambda_+)) = 1$. Namely, we only require

$$\begin{aligned} \exp(\lambda_-) &< \min\Big\{\frac{p(a)}{q(a)}; p(a) > 0, a \in A\Big\}, \\ \exp(\lambda_+) &> \max\Big\{\frac{p(a)}{q(a)}; q(a) > 0, a \in A\Big\}. \end{aligned} \tag{2.130}$$

As in (2.127), consider all the $\lambda_- \leq \lambda_1 \leq \cdots \leq \lambda_L \leq \lambda_+$ such that

$$\alpha(T_n^*(\lambda_+)) = \sum_{i=1}^{L} \Delta_n(\lambda_i) = 1, \tag{2.131}$$

then by assumption

$$0 < \Delta_n(\lambda_i) < \gamma, \qquad \text{for all } i \in \{1, \cdots, L\}. \tag{2.132}$$

Since $\alpha(T^*(\lambda_1)) = \Delta_n(\lambda_1) < \gamma$, we know there exists an $L_0 < L$, which is obviously dependent on n, such that

$$\sum_{i=1}^{L_0} \Delta_n(\lambda_i) \in (x_1, x_2). \tag{2.133}$$

Therefore take $\lambda = \lambda_{L_0}$ and then $\alpha(T^*(\lambda)) \in (x_1, x_2)$. What is left to show is that for any $\gamma > 0$ there exists an $M \in \mathbb{N}$ such that for all $n \geq M$ and all $\lambda \in \mathbb{R}$,

$$\Delta_n(\lambda) < \gamma. \tag{2.134}$$

Fix $\lambda \in \mathbb{R}$ and define the random variable, $Z : A \to \mathbb{R}$ as $Z(a) = \log p(a) - \log q(a)$ for $a \in A$ when $p(a) > 0$, and $Z(a) = 0$ otherwise, recall we always assume $p \ll q$. By construction $\mathbb{E}_p[Z] = D_{KL}(p||q)$, and if

$$S_n := Z_1 + \cdots + Z_n, \tag{2.135}$$

is the sum of n i.i.d copies of Z then,

$$\begin{aligned} p^n\big(S_n = n\lambda\big) &= p^n\Big(a \in A^n; \sum_{i=1}^{n} \log p(a_i) - \log q(a_i) = n\lambda\Big) \\ &= p^n\Big(a \in A^n; \log p^n(a) - \log q^n(a) = N\lambda\Big) \\ &= \Delta_n(\lambda). \end{aligned} \tag{2.136}$$

By the Central Limit Theorem A.2.6 if $\mu = \mathbb{E}[Z]$ and $\sigma = \mathbb{E}[|Z-\mu|^2]$ then for all $b > 0$ there exists an $M \in \mathbb{N}$ such that for all $n \geq M$,

$$\begin{aligned}
\Delta_n(\lambda) &= p^n\big(S_n = n\lambda\big) \\
&\leq p^n\left(\frac{1}{n}S_n - \lambda \in \left(\frac{-b\sigma}{\sqrt{n}}, \frac{b\sigma}{\sqrt{n}}\right)\right) \\
&\leq \sup_{\lambda\in\mathbb{R}} p^n\left(\frac{1}{n}S_n - \lambda \in \left(\frac{-b\sigma}{\sqrt{n}}, \frac{b\sigma}{\sqrt{n}}\right)\right) \\
&\leq \frac{\gamma}{2} + p^n\left(\frac{1}{n}S_n - \mu \in \left(\frac{-b\sigma}{\sqrt{n}}, \frac{b\sigma}{\sqrt{n}}\right)\right) \\
&< \frac{\gamma}{2} + \frac{1}{\sqrt{2\pi}}\int_{-b}^{b} \exp\big(-t^2\big)\mathrm{d}t. \qquad (2.137)
\end{aligned}$$

We can then pick b such that

$$\frac{1}{\sqrt{2\pi}}\int_{-b}^{b} \exp\big(-t^2\big)\mathrm{d}t < \frac{\gamma}{2}, \qquad (2.138)$$

and we can concludes the uniform convergence property holds in (2.128), and more importantly if we take $(x_1, x_2) = (\varepsilon - \delta, \varepsilon)$ then we can conclude (2.121).
Step 2: Define the set

$$\Omega_S := \big\{h \in \mathrm{Prob}(A); D_{KL}(h||q) > D_{KL}(h||p) + D_{KL}(p||q)\}, \qquad (2.139)$$

in this step we will show step 1 implies for all $\varepsilon > 0$ and $0 < \delta < \varepsilon$

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n}\log q^n\big(L_n \in \Omega_S\big) &\leq \liminf_{n\to\infty} \frac{1}{n}\log \beta^*_{n,\varepsilon-\delta}, \\
\limsup_{n\to\infty} \frac{1}{n}\log q^n\big(L_n \in \Omega_S\big) &\geq \limsup_{n\to\infty} \frac{1}{n}\log \beta^*_{n,\varepsilon-\delta}. \qquad (2.140)
\end{aligned}$$

Let $x \in A^n$ such that $p^n(x) > 0$ and with n_a we denote the number of times $a \in A$ appears in the string x, let $\lambda^* = D_{KL}(p||q)$ then the following three equations are equivalent,

$$\begin{aligned}
D_{KL}(L_n(x)||q) &> D_{KL}(L_n(x)||p) + \lambda^*, \\
-\sum_{a\in A} \frac{n_a}{n}\log q(a) &> -\sum_{a\in A} \frac{n_a}{n}\log p(a) + \lambda^*, \\
p^n(x) &> \exp(n\lambda^*)q^n(x), \qquad (2.141)
\end{aligned}$$

therefore,

$$\begin{aligned}
L_n^{-1}(\Omega_S) &= \big\{x \in A^n; D_{KL}(L_n(x)||q) > D_{KL}(L_n(x)||p) + \lambda^*\big\} \\
&= T_n^*(\lambda^*). \qquad (2.142)
\end{aligned}$$

Now recall the definition of the random variable S_n in (2.135). By (2.141) we then get the following two relations:

$$T_n^*(\lambda) = \{a \in A^n; S_n(a) > n\lambda\}, \qquad (2.143)$$

$$\alpha(T_n^*(\lambda)) = p^n(S_n \leq n\lambda). \qquad (2.144)$$

From the conclusion of step 1, namely (2.121), we know for all $n \in \mathbb{N}$ there exists a test $\lambda_n \in \mathbb{R}$ such that

$$\beta(T_n^*(\lambda_n)) < \beta_{n,\varepsilon-\delta}^* \left(1 + \frac{1}{n}\right). \tag{2.145}$$

Moreover, for all $n \in \mathbb{N}$ we require $\alpha(T_n^*(\lambda_n)) < \varepsilon$ and from the observations made in (2.130) we can assume $\lambda_n \in [\lambda_-, \lambda_+]$, therefore by compactness of the closed interval we can assume without loss of generality that $\lambda_n \to \lambda^*$ as $n \to \infty$ for some $\lambda^* \in [\lambda_-, \lambda_+]$. Now we will prove by contradiction that $\lambda^* = D_{KL}(p||q)$. By the weak law of large numbers, as stated in Theorem A.2.5, we have for all $\gamma > 0$,

$$\lim_{n\to\infty} p^n\Big(\Big|\frac{1}{n}S_n - D_{KL}(p||q)\Big| > \gamma\Big) = 0. \tag{2.146}$$

Suppose $\gamma = |D_{KL}(p||q) - \lambda^*| > 0$, then

$$\begin{aligned}
\lim_{n\to\infty} \alpha^*(T_n^*(\lambda^*)) &= \lim_{n\to\infty} p^n(S_n \le n\lambda^*) = 1 - \lim_{n\to\infty} p^n(S_n > n\lambda^*) \\
&= 1 - \lim_{n\to\infty} p^n\Big(\frac{1}{n}S_n - D_{KL}(p||q) > \lambda^* - D_{KL}(p||q)\Big) \\
&\ge 1 - \lim_{n\to\infty} p^n\Big(\Big|\frac{1}{n}S_n - D_{KL}(p||q)\Big| > \gamma\Big) \\
&= 1,
\end{aligned} \tag{2.147}$$

which is a contradiction since we require

$$\lim_{n\to\infty} \alpha(T_n^*(\lambda_n)) = \lim_{n\to\infty} \alpha(T_n^*(\lambda^*)) < \varepsilon. \tag{2.148}$$

Note, the the fact $\lim_{n\to\infty} \alpha(T_n^*(\lambda_n)) = \lim_{n\to\infty} \alpha(T_n^*(\lambda^*))$ is non-trivial and can can be shown by the inner and outer regularity measure $p^{\mathbb{N}}$. Putting everything together we get,

$$\liminf_{n\to\infty} \frac{1}{n} \log \beta(T_n^*(\lambda^*)) \le \liminf_{n\to\infty} \frac{1}{n} \log \beta_{n,\varepsilon-\delta}^*, \tag{2.149}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log \beta(T_n^*(\lambda^*)) \ge \limsup_{n\to\infty} \frac{1}{n} \log \beta_{n,\varepsilon-\delta}^*, \tag{2.150}$$

and using (2.142) we can conclude both equations in (2.140).

Step 3: In this step we use Sanov's theorem to prove Stein's lemma. We first treat the special setting that for all $a \in A$, $p(a) > 0$ and $q(a) > 0$. In this context the set Ω_S is open in $\mathrm{Prob}(A)$.[1] This follows from the fact that the function $F : \mathrm{Prob}(A) \to \mathbb{R}$ defined as

$$F(h) := D_{KL}(h||q) - D_{KL}(h||p), \tag{2.151}$$

is continuous, and

$$\Omega_S = F^{-1}\big((D_{KL}(p||q), \infty)\big). \tag{2.152}$$

It also easy to verify, for example by using sequences, that the closure of Ω_S is given by

$$\Omega_S^- = \{h \in \mathrm{Prob}(A); D_{KL}(h||q) \ge D_{KL}(h||p) + D_{KL}(p||q)\}. \tag{2.153}$$

[1] If we assume $p \ll q$ but there exists an $a \in A$ such that $q(a) > 0$ while $p(a) = 0$, then Ω_S is not open!

By Proposition II.2.3 we know

$$\inf_{h\in\Omega_S} D_{KL}(h||q) = \inf_{h\in\Omega_S^-} D_{KL}(h||q) = D_{KL}(p||q). \tag{2.154}$$

By Sanov's theorem and (2.140), for all $\varepsilon \in (0,1)$ and $0 < \delta < \varepsilon$, and $p, q \in \mathrm{Prob}(A)$ that have complete support on A, we find

$$\lim_{n\to\infty} \frac{1}{n} \log \beta^*_{n,\varepsilon-\delta} = -D_{KL}(p||q). \tag{2.155}$$

Now fix $\varepsilon \in (0,1)$ and pick $0 < \delta < \varepsilon$ such that $\varepsilon_0 = \varepsilon + \delta < 1$. Set $\delta_0 = \delta$, then $\varepsilon_0 - \delta_0 = \varepsilon$, and by (2.155) we conclude for all $p, q \in \mathrm{Prob}(A)$ that have complete support on A,

$$\lim_{n\to\infty} \frac{1}{n} \log \beta^*_{n,\varepsilon} = -D_{KL}(p||q). \tag{2.156}$$

To finish we will use to conclude Steins lemma for any two $p, q \in \mathrm{Prob}(A)$ such that only $p \ll q$, and $p \neq \delta_a$ for some $a \in A$. Let $A_0 := \{a \in A; p(a) > 0\}$; then $\mathrm{Prob}(A_0) \subseteq \mathrm{Prob}(A)$ is the set of all probability measures that are absolutely continuous with respect to p, which also means that $h \ll q$ for all $h \in \mathrm{Prob}(A_0)$. Now define $q_p \in \mathrm{Prob}(A_0)$ as follows. Let

$$Q = \sum_{a\in A_0} q(a), \tag{2.157}$$

and for all $a \in A_0$, define

$$q_p(a) := \frac{q(a)}{Q}. \tag{2.158}$$

We will explicitly write $\beta^*_{\varepsilon,n}(p, q_p)$ and $\beta^*_{\varepsilon,n}(p, q)$ to differentiate the two different hypothesis testing regimes, then applying (II.2.3) to $p, q_p \in \mathrm{Prob}(A_0)$ we obtain

$$\lim_{n\to\infty} \frac{1}{n} \log \beta^*_{\varepsilon,n}(p, q_p) = -D_{KL}(p||q_p). \tag{2.159}$$

Note in the case $q_p = p$ the above limit still holds since $D_{KL}(p||q_p) = 0$ and $\tilde{\beta}_{\varepsilon,N} \geq 1 - \varepsilon$. By the definition of q_p, we have

$$\beta^*_{\varepsilon,n}(p, q_p) = \frac{1}{q^n} \inf\{q^n(T_n); T_n \subseteq A_0^n; p^n(T_n) < \varepsilon\}. \tag{2.160}$$

Define

$$\tilde{\beta}_{\varepsilon,N} := \inf\{q^n(T_n); T_n \subseteq A_0^n; p^n(T_n) < \varepsilon\}, \tag{2.161}$$

and we get for all $n \in \mathbb{N}$,

$$\frac{1}{n} \log \beta^*_{\varepsilon,n}(p, q_p) = -\log Q + \frac{1}{n} \log \tilde{\beta}_{\varepsilon,N}, \tag{2.162}$$

therefore, since

$$D_{KL}(p||q_p) = D_{KL}(p||q) + \log(Q), \tag{2.163}$$

we then find the perhaps surprising relation

$$\lim_{n\to\infty} \frac{1}{n} \log \tilde{\beta}_{\varepsilon,n} = -D_{KL}(p||q). \tag{2.164}$$

Clearly, for all $n \in \mathbb{N}$, $\tilde{\beta}_{\varepsilon,N} \geq \beta_{n,\varepsilon}(p,q)$, while for all $\lambda \in \mathbb{R}$, $x \in T_n^*(\lambda) \implies p^n(x) > 0$, so

$$T_n^*(\lambda) \subset A_0^n \implies \tilde{\beta}_{\varepsilon,n} \leq \inf_{\lambda \in \mathbb{R}} \{\beta(T_n^*(\lambda)); \alpha(T_n^*(\lambda)) < \varepsilon\}. \tag{2.165}$$

By applying step 1 (2.121), we conclude Stein's lemma holds for all $p, q \in \text{Prob}(A)$ such that $p \ll q$ and $p \neq \delta_a$ for any $a \in A$. □

Proving Chernoff's lemma by applying Sanov theorem follows a very similar procedure as we have just completed for Stein's lemma. Below we provide an overview of the main steps of the proof of Chernoff's lemma.

Sketch Proof of Chernoff's Lemma II.2.3. Recall the definition of γ_n^* in (2.23), then by taking $\lambda = 0$ in Lemma II.2.6 we have

$$\gamma_n^* = \alpha(T_n^*(0)) + \beta(T_n^*(0)). \tag{2.166}$$

Define the set $\Omega_C \subseteq \text{Prob}(A)$ as,

$$\Omega_C := \big\{h \in \text{Prob}(A); D_{KL}(h||q) > D_{KL}(h||p)\big\}, \tag{2.167}$$

then taking $\lambda^* = 0$ in (2.142) we have

$$L_n^{-1}(\Omega_C) = T_n^*(0), \tag{2.168}$$

therefore

$$\alpha(T_n^*) = p^n(L_n \in \Omega_C^c), \quad \beta(T_n^*) = q^n(L_n \in \Omega_C), \tag{2.169}$$

where we write $T_n^* = T_n^*(0)$. If we assume for all $a \in A$, $p(a) > 0$ and $q(a) > 0$ then Ω_C is open and by Sanov's theorem we have

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \log p^n(L_n \in \Omega_C^c) &= -\inf_{h \in \Omega_C^c} D_{KL}(h||p) \\ \lim_{n\to\infty} \frac{1}{n} \log q^n(L_n \in \Omega_C) &= -\inf_{h \in \Omega_C} D_{KL}(h||q). \end{aligned} \tag{2.170}$$

Using the convexity of $(0,\infty) \ni x \mapsto \log x$ for all $n \in \mathbb{N}$ we have,

$$\begin{aligned} \frac{1}{n}\Big(\frac{1}{2}\log \alpha(T_n^*) + \frac{1}{2}\log \beta(T_n^*)\Big) &\leq \frac{1}{n}\Big(\log\frac{1}{2} + \log \gamma_n^*\Big) \\ &\leq \max\Big\{\frac{1}{n}\log\alpha(T_n^*), \frac{1}{n}\log\beta(T_n^*)\Big\}. \end{aligned} \tag{2.171}$$

Set $d_1 = \inf_{h\in\Omega_C^c} D_{KL}(h||p)$ and $d_2 = \inf_{h\in\Omega_C} D_{KL}(h||q))$ then we get

$$-\frac{1}{2}(d_1 + d_2) \leq \lim_{n\to\infty} \frac{1}{n} \log \gamma_n^* \leq \max\{-d_1, -d_2\}, \tag{2.172}$$

the final step involves proving $d_1 = d_2 = C(p,q)$ where the $C(p,q)$ is the Chernoff entropy defined in (2.26). We refer the reader to the [26] for the rest of the argument. □

Sketch proof of Theorem II.2.1. Let $\varepsilon > 0$ and $n \in \mathbb{N}$, we will prove the lower bound. Let $B, C \subseteq \text{Prob}(A)$, recall in the setting of compound hypothesis testing we defined the type I and II errors as

$$\beta(T_n) = \sup_{q \in C}\{q^n(T_n)\}, \quad \alpha(T_n) = \sup_{p \in B}\{p^n(T_n^c)\}, \tag{2.173}$$

and

$$\beta^*_{n,\varepsilon}(B, C) = \inf\{\beta(T_n); T_n \subseteq A^n, \alpha(T_n) < \varepsilon\}. \tag{2.174}$$

To prove the lower bound we will make use of two simple observations. Firstly for all $p \in B$ we have

$$\{T_n \subset A^n; \alpha(T_n) < \varepsilon\} \subseteq \{T_n \subseteq A^n; p^n(T_n^c) < \varepsilon\}, \tag{2.175}$$

and secondly for all $q \in C$ and $T_n \subseteq A^n$ we have $q^n(T_n) \leq \beta(T_n)$. Combining both of these observations we get for any two $p \in B$ and $q \in C$

$$\beta^*_{n,\varepsilon}(p, q) = \inf\{q^n(T_n); T_n \subseteq A^n, p^n(T_n^C) < \varepsilon\} \leq \beta^*_{n,\varepsilon}(B, C), \tag{2.176}$$

and by Stein's Lemma II.2.2,

$$-D_{KL}(p||q) = \lim_{n\to\infty} \frac{1}{n} \log \beta^*_{n,\varepsilon}(p, q) \leq \liminf_{n\to\infty} \frac{1}{n} \log \beta^*_{n,\varepsilon}(B, C), \tag{2.177}$$

and since this holds for $p \in B$ and $q \in C$, we have

$$\sup_{p\in B, q\in C}\{-D_{KL}(p||q)\} = -\inf_{p\in B, q\in C} D_{KL}(p||q) \leq \liminf_{n\to\infty} \frac{1}{n} \log \beta^*_{n,\varepsilon}(B, C). \tag{2.178}$$

We refer the reader to [8, Theorem 2] for a proof that

$$-\inf_{p\in B, q\in C} D_{KL}(p||q) \geq \limsup_{n\to\infty} \frac{1}{n} \log \beta^*_{n,\varepsilon}(B, C), \tag{2.179}$$

when B, C are both closed convex sets. Assuming $B, C \subseteq \text{Prob}(A)$ are both closed convex sets we conclude

$$-\inf_{p\in B, q\in C} D_{KL}(p||q) = \lim_{n\to\infty} \frac{1}{n} \log \beta^*_{n,\varepsilon}(B, C) \tag{2.180}$$

□

II.3 Sanov's theorem in more generality

In this section we state two generalizations of Sanov's theorem, where we specifically explain how the large deviation principle (Definition A.2.3) can be extended beyond the case $\mathcal{X} = \text{Prob}(A)$, where A is a finite set. The key building blocks for the Sanov's theorem are the following:

1. A measure space $(X, \mathcal{F})$,
2. the empirical measure $L_n : X^n \to \text{Prob}(X)$,
3. the relative entropy $D_{KL} : \text{Prob}(X) \times \text{Prob}(X) \to [0, \infty]$,
4. a topology τ' on $\text{Prob}(X)$,
5. and a $\sigma-$algebra $\mathcal{G}$ of subsets of $\text{Prob}(X)$.

The final point which requires a σ-algebra $\mathcal{G}$ can be omitted if $\mathcal{G}$ is the σ-algebra generated by the topology τ'. However, as we will soon see the Borel σ-algebra $\mathcal{B}_{\tau'}$ generated by a topology τ' on $\text{Prob}(X)$ is not always the correct choice of $\mathcal{G}$. We will first state a generalization of Sanov theorem to the setting where X is a Polish space. At a later stage in this thesis we use this result to prove a quantum Sanov theorem for a specific class of infinite dimensional von Neumann algebras. Luckily, in the Polish space setting we need only consider $\mathcal{G}$ as the Borel σ-algebra generated by the weak topology, in which case we say $\Omega \subseteq \text{Prob}(X)$ is Borel weakly measurable if $\Omega \in \mathcal{G}$.

Theorem II.3.1 (Sanov's theorem for Polish Spaces)**.** *Let X be a Polish space and $q \in \text{Prob}(X)$, with D_{KL} we denote the Kullback-Leibler relative entropy defined in* (II.1.6)*. Let $\Omega \in \mathcal{G}$ be Borel weakly measurable. Then*

$$
\begin{aligned}
-\inf_{p\in\Omega^\circ} D_{KL}(p||q) &\leq \liminf_{n\to\infty} \frac{1}{n} \log q^n(L_n \in \Omega) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \log q^n(L_n \in \Omega) \leq -\inf_{p\in\Omega^-} D_{KL}(p||q), \qquad (2.181)
\end{aligned}
$$

where Ω° is the interior of Ω and Ω^- the closure of Ω with respect to the weak topology.

Proof. See [12, Theorem 6.2.10]. □

Now we state the extension of Sanov's theorem to the case where X is a Hausdorff space, and even in the most general setting where X is a measure space. In the case that X is Polish, we are able to take the weak topology and then consider $\mathcal{G}$ to be the Borel weak σ-algebra. It is obvious that a different approach will be required when stating a Sanov's theorem when X is a general measure space, since there are no continuous bounded functions $f : X \to \mathbb{R}$ to construct the weak topology on $\text{Prob}(X)$. This is a reasonable motivation to consider the τ-topology on the set $\text{Prob}(X)$, since it only requires the measure space structure on X. However, simply taking $\mathcal{G}$ to be the the Borel σ-algebra generated by the τ-topology turns out to be insufficient. Notably, the empirical measure $L_n : X \to \text{Prob}(X)$ can fail to be a measurable function with respect to the product Borel σ-algebra generated by the τ-topology [15, 12]. Instead we need to define a smaller σ-algebra of subsets of $\text{Prob}(X)$. For an arbitrary measure space $(X, \mathcal{F})$ the cylinder

σ-algebra $\mathcal{G}_{cy}$ of $\mathrm{Prob}(X)$ is the smallest σ-algebra ensuring for all $f \in L^\infty(X)$ the maps $p_f : \mathrm{Prob}(X) \to \mathbb{R}$ defined as

$$p_f(p) := \int_X f \mathrm{d}p, \tag{2.182}$$

are measurable. Notice, that the τ-topology was defined as the smallest topology that requires the maps p_f to be continuous, but in general the Borel σ-algebra induced by the τ-topology need not equal the cylinder σ-algebra. Perhaps surprisingly, the discussion becomes more interesting when we consider X as a Hausdorff space, since in this case we are able to define the weak topology, as well as the and τ-topology, and hence there are two Borel σ-algebras that seem reasonable. It turns out the only viable choice is taking $\mathcal{G} = \mathcal{G}_{cy}$.

This discussion will become especially relevant in the last chapter of this thesis. There we investigate how we can extend a *quantum Sanov theorem* in the finite dimensional von Neumann algebra setting to the infinite dimensional von Neumann algebra setting. We already point out that the transition from finite to infinite dimensional von Neumann algebras in the quantum setting shares similarities to the transition from $X = A$ being a finite set to X being a Polish space, Hausdorff space, or an arbitrary measure space in the classical setting. To finish this chapter we state the Sanov's theorem for when X is a general measure space, therefore also including the case that X is Hausdorff.

Theorem II.3.2 (Sanov's theorem for measure spaces)**.** *Let $(X, \mathcal{F})$ be a measure space and $q \in \mathrm{Prob}(X)$. With D_{KL} we denote the Kullback-Leibler relative entropy defined in* (II.1.6)*. Let $\Omega \in \mathcal{G}_{cy}$. Then*

$$\begin{aligned} -\inf_{p\in\Omega^\circ} D_{KL}(p||q) &\leq \liminf_{n\to\infty} \frac{1}{n} \log q^n(L_n \in \Omega) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log q^n(L_n \in \Omega) \leq -\inf_{p\in\Omega^-} D_{KL}(p||q), \end{aligned} \tag{2.183}$$

where Ω° is the interior of Ω and Ω^- the closure of Ω with respect to the $\tau-$topology.

Proof. See [1, Theorem 1.1]. □

If we return to the case that X is Polish, then the Borel σ-algebra generated by the weak topology equals the cylinder σ-algebra $\mathcal{G}_{cy}$.[2] Despite there being no ambiguity in the choice of the σ-algebra $\mathcal{G}$, if we compare Theorem II.3.1 and Theorem II.3.2 then the remaining difference is that for $\Omega \in \mathcal{G}_{cy}$, the closure (or interior) of Ω is different for the τ-topology and the weak topology, therefore these two results remain technically distinct.

[2] See Exercise 6.2.18 in [12].

Chapter III

Quantum probability

We recall the main objective of this thesis is to gain a better understanding of randomness in quantum mechanics. Our method is to investigate the connections between classical and quantum probability theory with a key focus on Sanov's theorem. In the setting of probability theory, we view Sanov's theorem as a foundational result which *describes* the interaction we as observers have with (classical) random systems. If we limit our interactions with random systems to sampling i.i.d events then, the empirical measure models our interaction and, the description of that interaction is given by the asymptotic bounds in Sanov's theorem. In this chapter we state a few proposed quantum Sanov theorem's along with the quantum extensions of Stein's and Chernoff's lemmas. We will follow a similar structure as in Chapter II. We will provide a full prove of the quantum Stein lemma and of the first proposed quantum Sanov theorem. When proving the main results in this section we restrict ourselves to the setting of finite dimensional quantum probability, and in Chapter IV we will treat the infinite dimensional case. In Section III.1 we start off with a preliminary section and we stablish the foundational ideas in quantum probability. To illustrate that quantum probability is truly a non-commutative extension of classical probability, in Section III.2 we restate the Sanov Theorem II.2.2 in a commutative algebraic language.

In Section III.3.1, we discuss the mathematical framework of measuring/sampling finite dimensional quantum states. This topic is particularly nuanced since there is no classical intuition to build on. In Section III.3.2 we cover the basics of quantum hypothesis testing, which shares remarkably similar definitions to the classical case. The rest of this chapter serves as a detailed literature review of the very active field of quantum probability theory. We first prove the quantum Stein lemma in Section III.3.3. Directly afterwards we will apply the quantum Stein lemma to partially prove the first quantum extension of Sanov's theorem proposed in [6], which we view as a quantum extension of the Operational Sanov Theorem II.2.1. Afterwards, we will introduce a result of a very recent paper [16], titled *Another quantum Sanov Theorem*, which makes use of Schur-Weyl duality to address a shortcoming of the first quantum Sanov theorem proposed in [6]. Understanding both of these papers is central for the rest of our thesis, since we offer an alternative quantum Sanov theorem that builds on the ideas presented in [6, 16].

III.1 Preliminaries

In this section we introduce the foundations of quantum probability theory. We assume the reader has no prior exposure to quantum mechanics, thereby offering the alternative title of noncommutative probability theory for this chapter. We assume the reader is familiar with basics of functional analysis, however we still state the essential definitions and results in this section. Note in Appendix B we provide a generic overview of von Neumann algebras and functional analysis as a refresher. As opposed to the rest of this chapter, we will not restrict ourselves to the finite dimensional quantum systems and instead work define concepts in as much generality as possible. However, besides for defining normal states in arbitrary von Neumann algebras (Definition B.0.9) we will mostly be working with type I_d factored von Neumann algebras for $d \in \mathbb{N}$. The complete classification of a factored von Neumann algebras will be discussed at a later stage in this thesis.

Definition III.1.1 (Type I von Neumann Algebra)**.** *A von Neumann algebra $\mathcal{M}$ is a **type I factor** if there exists a complex Hilbert space $\mathcal{H}$ such that $\mathcal{M}$ is algebraically isomorphic (Definition B.0.11) to $B(\mathcal{H})$, i.e. $\mathcal{M} \simeq B(\mathcal{H})$. Let $d \in \mathbb{N}$. We call $\mathcal{M}$ a type I_d factor if $\mathcal{M} \simeq B(\mathcal{H})$ and $\mathcal{H}$ is a d-dimensional Hilbert space, and a type I_∞ factor if $\mathcal{M} \simeq B(\mathcal{H})$ and $\mathcal{H}$ is an infinite dimensional Hilbert space.*

Because a von Neumann algebra $\mathcal{M} \subseteq B(\mathcal{H})$ is always a subset of some larger space of bounded operators $B(\mathcal{H})$, we are able to capitalize on the structure of $B(\mathcal{H})$ when working with $\mathcal{M}$. Through the rest of this thesis we will use $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ to denote the inner product on the Hilbert space $\mathcal{H}$. With $|| \cdot || : \mathcal{H} \to \mathbb{R}^+$ we denote the norm $||x|| := \sqrt{\langle x, x \rangle}$ on $\mathcal{H}$. When convenient, for a unit vector $e \in \mathcal{H}$ we will use bra-ket notation and write $|e\rangle \in \mathcal{H}$ to denote the vector, and $|e\rangle \langle e| \in B(\mathcal{H})$ to denote the one dimensional projection onto the linear space spanned by e. The following class of operators in $B(\mathcal{H})$ will be very useful.

Definition III.1.2 (Trace class operators $L^1(\mathcal{H})$)**.** *The set of **trace class operators** $L^1(\mathcal{H}) \subset B(\mathcal{H})$ is defined as*

$$L^1(\mathcal{H}) := \left\{ a \in B(\mathcal{H}); \sum_{e \in E} \langle e, |a| e \rangle < \infty, \text{ for an orthonormal basis E of } \mathcal{H} \right\}, \tag{3.1}$$

where $|a| = \sqrt{aa^*} \in B(\mathcal{H})$ is defined by the functional calculus on $B(\mathcal{H})$ (Theorem B.0.4). Note that the definition is independent of the orthonormal basis E.

For all trace class operators we are then able to define the trace of an operator.

Definition III.1.3 (Trace)**.** *For any Hilbert space $\mathcal{H}$ we can define the **trace** $\mathrm{Tr} : L_1(\mathcal{H}) \to \mathbb{C}$ as*

$$\mathrm{Tr}[v] = \sum_{e \in E} \langle ve, e \rangle, \tag{3.2}$$

for one and hence any orthonormal basis E.

On the set $B(\mathcal{H})$ we are able to define several topologies, the two we will most frequently use are the operator or uniform topology and the σ-weak topology.

Definition III.1.4 (Topologies in $B(\mathcal{H})$)**.** *Let $\mathcal{H}$ be a Hilbert space and $B(\mathcal{H})$ be the space of bounded linear operators.*

1. *The **uniform/norm topology** is generated by the operator norm $||\cdot||_{op} : B(\mathcal{H}) \to \mathbb{R}^+$, defined as*

$$||a||_{op} = \sup\{||ax||; x \in \mathcal{H}, ||x|| = 1\}. \tag{3.3}$$

2. *The **σ-weak topology** is the locally convex topology on $B(\mathcal{H})$ defined by the family of semi-norms $(p_v)_{v \in L^1(\mathcal{H})}$ defined as*

$$p_v(a) = |\mathrm{Tr}[va]|. \tag{3.4}$$

Equivalently, a net of operators $(a_\lambda)_{\lambda\in\Lambda}$ converges $\sigma-$weakly to $a_0 \in B(\mathcal{H})$ if for all $v \in L_1(\mathcal{H})$ the net $(p_v(a_\lambda))_{\lambda\in\Lambda}$ converges to $p_v(a_0)$ in $\mathbb{R}^+$.

The most fundamental definition in probability theory we need to extend to a noncommutative alternative is that of a probability measure $p \in \mathrm{Prob}(\Omega)$ over some measure space $(\Omega, \mathcal{F})$. Here we will give the very general extension of probability measures to quantum probability. We will make use of the following notation. Fix an arbitrary von Neumann algebra $\mathcal{M} \subseteq B(\mathcal{H})$ and let $\mathbf{1} \in \mathcal{M}$ represent the identity operator. With $\mathcal{M}^*$ we denote the Banach dual (Definition B.0.2) of $\mathcal{M}$, i.e. the space of all norm continuous linear functionals $\psi : \mathcal{M} \to \mathbb{C}$. Let $\mathcal{M}^+$ represent all positive operators (Definition B.0.7) in $\mathcal{M}$. We can now define a state on a von Neumann algebra.

Definition III.1.5 (States)**.** *We call a functional $\psi \in \mathcal{M}^*$ a **state** if and only if $\psi(\mathbf{1}) = 1$, and for all positive operators $a \in \mathcal{M}^+$, $\psi(a) \geq 0$. The set of states is denoted as $\mathcal{S}(\mathcal{M})$.*

The noncommutative alternative to $\mathrm{Prob}(\Omega)$ is then defined as the set of all normal states $\mathcal{S}_n(\mathcal{M})$ on $\mathcal{M}$.

Definition III.1.6. *a functional $\psi \in \mathcal{M}^*$ is called **normal** if is is continuous with respect to the σ-weak topology. The set of all normal functionals on $\mathcal{M}$ is called the **predual**, and it is denoted as $\mathcal{M}_*$.*

We have the following useful characterization of normal states.

Proposition III.1.1. *Let $\mathcal{M} \subseteq B(\mathcal{H})$ be an arbitrary von Neumann algebra and $\varphi : \mathcal{M} \to \mathbb{C}$ a state. Then φ is $\sigma-$weakly continuous if and only there exists a $\rho \in L_1(\mathcal{H})$ with $\mathrm{Tr}[\rho] = 1$ and $\rho \geq 0$, such that for all $a \in \mathcal{M}$*

$$\varphi(a) = \mathrm{Tr}[\rho a]. \tag{3.5}$$

Proof. See [29, Theorem 4.2.10]. □

For an arbitrary Hilbert space $\mathcal{H}$ we can now define the set of density operators $D(\mathcal{H})$, this set will turn out to be a more useful noncommutative extension of $\mathrm{Prob}(\Omega)$, when compared to $\mathcal{S}_n(\mathcal{M})$. However, it comes at a cost.

Definition III.1.7 (Density operators)**.** *Let $\mathcal{H}$ be an arbitrary Hilbert space, then the set of* ***density operators*** *$D(\mathcal{H}) \subset B(\mathcal{H})$ is defined as*

$$D(\mathcal{H}) := \big\{\rho \in L^1(\mathcal{H}); \mathrm{Tr}(\rho) = 1, \rho \geq 0\big\}. \tag{3.6}$$

Notice, for all $\rho \in D(\mathcal{H})$, $\rho = \rho^*$ is self adjoint.

For an arbitrary von Neumann algebra $\mathcal{M}$ any normal state $\psi \in \mathcal{S}_n(\mathcal{M})$ can be associated to a density operator $\rho \in D(\mathcal{H})$ due to Proposition III.1.1. However, a crucial point to notice is that unless $\mathcal{M} = B(\mathcal{H})$, ρ is not guaranteed to be unique.

Proposition III.1.2. *Suppose $\mathcal{M} = B(\mathcal{H})$ is a type I factor and $\mathcal{H}$ is separable. Then for a normal state $\psi \in \mathcal{S}_n(\mathcal{M})$ the density operator obtained in Proposition III.1.1 is unique.*

Proof. Suppose there are two density operators $\rho_1, \rho_2 \in D(\mathcal{H})$ such that $\rho_1 \neq \rho_2$ but for all $a \in B(\mathcal{H})$ we have

$$\mathrm{Tr}[(\rho_1 - \rho_2)a] = 0. \tag{3.7}$$

Because $\rho_1 - \rho_2$ is a self adjoint we know by the spectral decomposition Theorem B.0.3, there exists a unit vector $|\phi\rangle \in \mathcal{H}$ such that $\langle \phi, (\rho_1 - \rho_2)\phi\rangle \neq 0$, then take $a = |\phi\rangle\langle\phi|$ and

$$\mathrm{Tr}[(\rho_1 - \rho_2)a] = \langle \psi, (\rho_1 - \rho_2)\psi\rangle \neq 0. \tag{3.8}$$

□

As portrayed in the simple proof above, since density operators $\rho \in B(\mathcal{H})$ are bounded, self adjoint, and trace class, there is a great amount of tools available to manipulate them. In the case that $\mathcal{M} \subset B(\mathcal{H})$ is a strict subset, then it can turn out that there are two density operators $\rho_1 \neq \rho_2$ but for all $a \in \mathcal{M}$ we have

$$\mathrm{Tr}[(\rho_1 - \rho_2)a] = 0. \tag{3.9}$$

When a density operator associated to a normal state is not unique, then techniques involving Tomita-Takesaki theory are required to define even the most fundamental definitions for normal states, such as the quantum variant of the Kullback-Leibler relative entropy. For the rest of this section we will therefore restrict ourselves to studying von Neumann algebras $\mathcal{M} = B(\mathcal{H})$ which are type I factors. This means that instead of considering $\mathcal{S}_n(\mathcal{M})$ as the noncommutative alternative of $\mathrm{Prob}(\Omega)$, we can equivalently consider the set of density operators $D(\mathcal{H})$. It is standard practice in quantum information theory to slightly abuse our definitions and call density operators $\rho \in D(\mathcal{H})$ quantum states. In the situation $\mathcal{M} = B(\mathcal{H})$ by Proposition III.1.2, $D(\mathcal{H})$ is in bijection with $\mathcal{S}_n(\mathcal{M})$. Namely, for a normal state $\varphi \in \mathcal{S}_n(\mathcal{M})$ there exists a unique $\rho \in D(\mathcal{H})$ such that $\varphi(a) = \mathrm{Tr}[\rho a]$ for all $a \in \mathcal{M}$. This duality enables us to refer to both ψ and ρ as quantum states.

We also wish to point out that in the classical probability setting, we defined several topologies, such as the weak topology and τ-topology, on the set $\mathrm{Prob}(\Omega)$. The advantage of working on $D(\mathcal{H})$ is that we can inherent any topology in the larger space $B(\mathcal{H})$.

Definition III.1.8 (Umegaki's relative entropy)**.** *Let $\rho, \sigma \in D(\mathcal{H})$ be normal states on $\mathcal{M} = B(\mathcal{H})$. Then the* ***Umegaki relative entropy*** *$D(\cdot||\cdot) : D(\mathcal{H}) \times D(\mathcal{H}) \to [0, \infty]$ is defined as*

$$D(\rho||\sigma) := \begin{cases} \mathrm{Tr}[\rho(\log \rho - \log \sigma)] & \text{if } \ker(\sigma) \subseteq \ker(\rho) \\ \infty, & \text{otherwise.} \end{cases} \tag{3.10}$$

We are also able to define a quantum extension of the Rényi relative entropy.

Definition III.1.9 (Petz-Rényi Relative Entropy)**.** *Let $\alpha \in \mathbb{R}\backslash\{0, 1\}$ then the α-**Petz-Rényi relative entropy** $R_\alpha : D(\mathcal{H}) \times D(\mathcal{H}) \to [0, \infty]$ is defined as*

$$R_\alpha(\rho||\sigma) := \begin{cases} \frac{1}{\alpha - 1} \log \mathrm{Tr}[\rho^\alpha \sigma^{1-\alpha}]. & \text{if } \ker(\sigma) \subseteq \ker(\rho) \\ \infty, & \text{otherwise.} \end{cases} \tag{3.11}$$

We define the quantum extensions of the Kullback Leibler relative entropy and the Rényi relative entropy for density operators, however it is possible to extend the definition to arbitrary normal states $\psi, \phi \in \mathcal{S}_n(\mathcal{M})$. The reason we do not present these definition here is mostly that there is no existing literature which has proven a quantum version of Sanov's theorem beyond the finite dimensional type I_d case.[1] In fact, from the best of our knowledge the universal quantum Sanov theorem we prove later on is the most general quantum version of Sanov theorem; as we prove it in the type I_∞ case. We can now state a collection of useful and well known properties of the Umegaki and and Petz-Rènyi relative entropies.

Definition III.1.10 (CPTP Map)**.** *Let $\mathcal{H}$ and $\mathcal{K}$ be two separable Hilbert spaces. A linear and bounded map $\Phi : L^1(\mathcal{H}) \to L^1(\mathcal{K})$ is* ***trace preserving*** *if for all $a \in L^1(\mathcal{H})$ we have that*

$$\mathrm{Tr}[\Phi(a)] = \mathrm{Tr}[a]. \tag{3.12}$$

The map Φ is ***positive*** *if for all $a \in L^1(\mathcal{H})$ satisfying $a \geq 0$ then $\Phi(a) \geq 0$. Let $M_k(\mathbb{C})$ be the k by k complex matrices, furthermore let $\boldsymbol{I}_k : M_k(\mathbb{C}) \to M_k(\mathbb{C})$ be the identity map for all $k \in \mathbb{N}$. We say Φ is* ***completely positive*** *if for all $k \in \mathbb{N}$ the map*

$$\Phi \otimes \boldsymbol{I}_k : B(\mathcal{H}) \otimes M_k(\mathbb{C}) \to B(\mathcal{H}) \otimes M_k(\mathbb{C}), \tag{3.13}$$

is positive. A map that is completely positive and trace preserving is called a ***CPTP*** *map.*

Proposition III.1.3. *Let $\mathcal{H}$ be a separable Hilbert space and $\rho, \sigma \in D(\mathcal{H})$.*

1. *$D(\rho||\sigma) \geq 0$ and $D(\rho||\sigma) = 0$ if and only if $p = q$.*
2. *The map $\rho, \sigma \mapsto D(\rho||\sigma)$ is jointly convex.*
3. *(Monotonicity) For all trace preserving completely preserving positive maps $\Phi : L^1(\mathcal{H}) \to L^1(\mathcal{H})$, we have the*

$$D(\rho||\sigma) \geq D(\Phi(\rho)||\Phi(\sigma)). \tag{3.14}$$

[1] On the contrary, Stein's lemma has successfully been developed well beyond the type I_d case.

Proof. For the first two points see [19, Theorem 2.7] for a much more general statement. For the monotonicity see [27]. □

Proposition III.1.4. *Let $\mathcal{H}$ be a separable Hilbert space and $\rho, \sigma \in D(\mathcal{H})$. Then for all $\alpha \in \mathbb{R}\backslash\{0,1\}$,*

1. *$R_\alpha(\rho||\sigma) \geq 0$.*
2. *The map $\rho, \sigma \mapsto R_\alpha(\rho||\sigma)$ is jointly convex.*

Proof. See [19, Theorem 3.2] for a much more general statement. □

III.2 The quantum formalism of classical probability

It is insightful to consider that one can completely describe the classical setting of finite probability spaces in terms of operators. To this end, we will show how there are algebraic counterparts to the set of probability measures $\mathrm{Prob}(A)$ and, the σ-algebra $\{T; T \subseteq A\}$ of the events. Fix a $d \in \mathbb{N}$ and assume $A = \{1, \cdots, d\}$. Let $\mathcal{H}$ be a d-dimensional Hilbert space with an orthonormal basis $\mathcal{B} = (|x\rangle)_{i=1}^d$. Trivially we have a one-to-one correspondence between A and $\mathcal{B} = (|a\rangle)_{a\in A}$. In fact, we can take any probability measure $p \in \mathrm{Prob}(A)$ and define

$$\rho_p := \sum_{a\in A} p(a) \, |a\rangle \langle a| \in D(\mathcal{H}). \tag{3.15}$$

On the other hand, for any density operator ρ of the form

$$\rho = \sum_{a\in A} \lambda_a \, |a\rangle \langle a| \,, \tag{3.16}$$

we can associate a probability distribution $p_\rho \in \mathrm{Prob}(A)$ defined as $p_\rho(a) := \lambda_a$, for $a \in A$. Naturally, for a the fixed basis $\mathcal{B} = \{|a\rangle\}_{a\in A}$, normal states $\sigma \in D(\mathcal{H})$ are not always diagonalizable with respect to $\mathcal{B}$, i.e. of the form (3.16). To collect all normal states which are diagonalizable with respect to $\mathcal{B}$, we define $D_\mathcal{B}(\mathcal{H}) \subseteq D(\mathcal{H})$ as

$$D_\mathcal{B}(\mathcal{H}) := \{\rho \in \mathcal{D}(\mathcal{H}); \rho = \sum_{a\in A} \lambda_a \, |a\rangle \langle a| \text{ for a convex combination } (\lambda_a)_{a\in A}\}. \tag{3.17}$$

We can now state the quantum formalism of (finite) probability theory.

Proposition III.2.1. *The function $\Phi_\mathcal{B} : \mathrm{Prob}(A) \to D_\mathcal{B}(\mathcal{H})$ defined as*

$$\Phi_\mathcal{B}(p) = \rho_p,$$

with inverse $\Phi^{-1}(\rho) = p_\rho$, is bijective and continuous with respect to the Euclidean topology on both sets. Namely, on $\mathrm{Prob}(A)$ and $D_\mathcal{B}(\mathcal{H})$ we use the metrics d_1 and d_2 defined as,

$$d_1(p,q) = \sqrt{\sum_{a\in A} |p(a) - q(a)|^2} \quad \textit{for } p, q \in \mathrm{Prob}(A), \tag{3.18}$$

$$d_2(\rho_p, \rho_q) = \sqrt{\sum_{a\in A} |p(a) - q(a)|^2} \quad \textit{for } \rho_p, \rho_q \in D_\mathcal{B}(\mathcal{H}). \tag{3.19}$$

Proof. Take an arbitrary $\rho \in D_{\mathcal{B}}(\mathcal{H})$, then $\rho = \sum_{a\in A} \lambda_a |a\rangle \langle a|$, where $\lambda_a \geq 0$ and $\sum_{a\in A} \lambda_a = 1$. Define the probability distribution $p_\rho \in \text{Prob}(A)$ as $p_\rho(a) = \lambda_a$ for $a \in A$, and $\Phi(p_\rho) = \rho$, so Φ is are surjective. To prove Φ is injective let $p, q \in \text{Prob}(A)$ such that $\Phi(p) = \rho_p = \rho_q = \Pi(q)$, then $\rho_p - \rho_q = 0$ and this is true only if for all $a \in A$,

$$\langle a| \rho_p |a\rangle = p(a) = q(a) = \langle a| \rho_q |a\rangle , \tag{3.20}$$

and Φ is injective. The continuity of Φ follows from observing if $p_n \to p$ with respect to d_1, then

$$\lim_{n\to\infty} d_2(\Phi_{\mathcal{B}}(p_n), \Phi_{\mathcal{B}}(p)) = \lim_{n\to\infty} d_2(\rho_{p_n}, \rho_p) = \lim_{n\to\infty} d_1(p_n, p) = 0. \tag{3.21}$$

□

We will now discuss how the quantum formalization of an event $E \subseteq A$ can be described by some projection $P_E \in B(\mathcal{H})$. Let $q \in \text{Prob}(A)$ and $M = \{a_1, \cdots, a_m\} \subseteq A$. Define the projection

$$P_M = \sum_{a\in M} |a\rangle \langle a| , \tag{3.22}$$

which we can use to express the probability of the event M occurring with respect to the measure q as

$$\begin{aligned} q(M) &= \sum_{a\in M} q(a) = \sum_{x\in A} \sum_{a\in M} \delta_{xa} q(x) \\ &= \sum_{x\in A} q(x) \langle x| P_M |x\rangle = \sum_{x\in A} \langle x| P_M \Big(\sum_{a\in A} q(a) |a\rangle \langle a| \Big) |x\rangle \\ &= \text{Tr}[P_M \rho_q]. \end{aligned} \tag{3.23}$$

An essential concept in quantum probability theory is that the projection P_M should be interpreted as an *event* that can occur when one *measures* a quantum state $\rho_q \in D_{\mathcal{B}}(\mathcal{H})$. We will discuss the process of quantum measurements in greater detail in the next section but, the key takeaway is that $\text{Tr}[\rho_q P_M] \in [0, 1]$ is the probability that one *measures* the event P_M when the quantum state is ρ_q.

Moving on, we can extend the notion of a Bernoulli measure q^n on A^n to the operator $\rho_q^{\otimes n}$ in $B(\mathcal{H}^{\otimes n})$ since for a string $x = (x_1, \cdots x_n) \in A^n$ if $P_{x_i} \in B(\mathcal{H})$ is a projection on the span of $|x_i\rangle$ for $i \in \{1, \cdots, N\}$ then

$$\begin{aligned} q^n(\{x_1, \cdots, x_n\}) = q(x_1) \cdots q(x_n) &= \prod_{j=1}^{n} \text{Tr}[P_{x_i} \rho_p] \\ &= \text{Tr}\Big[\big(P_{x_1} \otimes \cdots \otimes P_{x_n} \big) \cdot \rho_q^{\otimes n} \Big]. \end{aligned} \tag{3.24}$$

We extend this to arbitrary sets $M_n \subseteq A^n$ simply by linearity,

$$P(M_n) := \sum_{x\in M^n} P_{x_1} \otimes \cdots \otimes P_{x_n}. \tag{3.25}$$

Note, the tensor product of n projections is also a projection and, for two $x, y \in A^n$, we have $\big(P_{x_1} \otimes \cdots \otimes P_{x_n}\big)\big(P_{y_1} \otimes \cdots \otimes P_{y_n}\big) = 1$ if and only if $x = y$, and equals zero otherwise. We can confirm $\rho_q^{\otimes n}$ is a sufficient quantum variant on the n-fold Bernoulli measure of p as follows,

$$\mathrm{Tr}[\rho_p^{\otimes n} P_{M_n}] = \sum_{a \in M_n} \mathrm{Tr}[\rho_p P_{a_1}] \cdots \mathrm{Tr}[\rho_p P_{a_n}] = p^n(M_n). \tag{3.26}$$

Finally, we can restate Sanov's theorem in the language of algebraic quantum mechanics:

Theorem III.2.1 (Quantum Formalism of the Classical Sanov Theorem)**.** *Let $\sigma \in \mathcal{D}_{\mathcal{B}}(\mathcal{A})$ and recall the function Φ defined in Proposition III.2.1. Let $\mathcal{Q} \subseteq \mathcal{D}_{\mathcal{B}}(\mathcal{A})$ such that $\Phi^{-1}(\mathcal{Q}) = Q \subseteq \mathrm{Prob}(A)$ is measurable. For $N \in \mathbb{N}$, define the projection associated to the empirical measure $P_n = P_{L_n^{-1}(Q)} \in B(\mathcal{H}^{\otimes n})$. Then*

$$\begin{aligned} -\inf_{\rho \in \mathcal{Q}^\circ} D(\rho||\sigma) \leq & \liminf_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} P_n] \\ & \leq \limsup_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} P_n] \leq -\inf_{\rho \in \mathcal{Q}^-} D(\rho||\sigma), \end{aligned} \tag{3.27}$$

where $\mathcal{Q}^\circ$ and $\mathcal{Q}^-$ is the interior and closure of $\mathcal{Q}$ with respect to topology induced by the d_2 metric (3.19)*.*

Proof. Let $q \in \mathrm{Prob}(A)$ be the unique probability distribution such that $\Phi(q) = \sigma$. Since $P_n = P(L_n^{-1}(\Phi^{-1}(\mathcal{Q})))$ by (3.26) we have $\mathrm{Tr}[\sigma P_n] = q^n(L_n^{-1}(Q))$. Furthermore take any two $\rho_1, \rho_2 \in \mathcal{Q} \subset \mathcal{D}_{\mathcal{B}}(\mathcal{H})$, and $p_1, p_2 \in \mathrm{Prob}(A)$ such that $\Phi(p_i) = \rho_i$ for $i = 1, 2$. Then $p_1 \ll p_2$ if and only if $\ker(\rho_1) \subseteq \ker(\rho_2)$ and

$$\begin{aligned} D_{KL}(p_1||p_2) &= -\sum_{a \in A} p_1(a)(\log p_1(a) - \log q_1(a)) \\ &= \sum_{a \in A} \langle a| \Big(\sum_{a' \in A} p_1(a')(\log p_1(a') - \log p_2(a')) \, |a'\rangle \langle a'| \Big) |a\rangle \\ &= -\mathrm{Tr}[\rho_1(\log \rho_1 - \log \rho_2)] = D(\rho_1||\rho_2). \end{aligned} \tag{3.28}$$

Since Φ is bijective and continuous we know

$$\inf_{\rho_1 \in \mathcal{Q}} D(\rho_1||\sigma) = \inf_{p_1 \in Q} D_{KL}(p_1||q), \tag{3.29}$$

and the statement follows from the classical Sanov Theorem II.2.2. □

III.3 Quantum measurement and hypothesis testing

III.3.1 Measuring quantum states

The procedure of measuring a quantum state differs greatly from measuring or sampling a classical probability distributing. To clarify this point considering the following finite dimensional comparison. Let A be a finite set such that $|A| = d$ and let $\mathcal{H}$ be a d-dimensional Hilbert space. Fix a probability distribution $p \in \text{Prob}(A)$ and a quantum state $\rho \in D(\mathcal{H})$. In the classical case suppose X is a random variable taking values on A with law p. The random variable X is then the mathematical representation of a measurable *property* of some random system, where the properties of the system are different elements in the set A. Simple examples include: X representing the (floating) exchange rate of a currency in a free market, or the speed of the wind at a weather station. Throughout the previous chapter on classical probability theory we emphasized that the empirical measure is a model for how we as observers interact with classical randomness. Going to the axioms of Kolmogorov's probability theory the meaning of the probability of an event is to be understood as the limit of the frequency one would measure the event, upon repeatedly sampling the same random variable X.

From our current understand of quantum mechanics, there is no intrinsic method of sampling a quantum state $\rho \in D(\mathcal{H})$ as there is for sampling a probability distribution $p \in \text{Prob}(A)$. To make sense of this we must first introduce a few assumption we need to make about how we can physically interact with quantum states. Firstly, as per the foundations of quantum mechanics we assume that one is always able to device some physical experiment which allows one to measure a fixed self adjoint operator $a \in B(\mathcal{H})$. Since $\mathcal{H}$ is finite dimensional, by the Spectral Theorem B.0.3 we can express the self adjoint operator a as follows

$$a = \sum_{i=1}^{d} \lambda_i |i\rangle \langle i| , \tag{3.30}$$

where $\{|i\rangle\}_{i=1}^{d}$ is an orthonormal basis and $\lambda_1, \cdots, \lambda_d$ are the eigenvalues of a. Our second assumption is Born's rule, which offers a conceptual connection from quantum states to classical probability distributions.

Definition III.3.1 (Born's rule)**.** *For a normal state $\rho \in D(\mathcal{H})$ (or $\psi \in \mathcal{S}(\mathcal{M})$), if one measures a self adjoint operator $a = a^*$ then one will sample an eigen value $\lambda \in \{\lambda_1, \cdots, \lambda_d\}$ of a. Moreover, if can repeatedly reproduce the quantum state ρ and measure the self adjoint operator $a = a^*$, then the expected value of the eigenvalue measurement is given by* $\text{Tr}[\rho a]$.

It is useful to consider how Born's rule applies to the quantum equivalent of a Dirac distribution $\delta_a \in \text{Prob}(A)$. We call ρ a **pure** normal state, if $\rho = |\psi\rangle \langle\psi|$ for some unit vector $|\psi\rangle \in \mathcal{H}$. Note the common feature between $\delta_a \in \text{Prob}(A)$, and $\rho = |\psi\rangle \langle\psi| \in D(\mathcal{H})$, is that both are *extreme points* of the respective convex sets $\text{Prob}(A)$ and $D(\mathcal{H})$. By Born's rule we can consider the measure of the (self adjoint) projection $|\psi\rangle \langle\psi|$, which has eigenvalues 0 and 1. Born's rule then states the expected eigenvalue one will obtain by measuring ρ with $|\psi\rangle \langle\psi|$ is given by $\text{Tr}[\rho |\psi\rangle \langle\psi|] = 1$. Therefore, much like in the classical case, we are certain of the outcome of measuring ρ with the projection $|\psi\rangle \langle\psi|$,i.e. we will always measure 1. The interesting difference between the classical Dirac distribution and

quantum pure state $\rho = |\psi\rangle\langle\psi|$, is that there are projections (or events), $|\gamma\rangle\langle\gamma| \in B(\mathcal{H})$ such that $\mathrm{Tr}[|\gamma\rangle\langle\gamma|\,\rho] \in (0,1)$, and therefore the pure normal states are not deterministic. This is of course very unnatural in the classical setting, since the Dirac distribution δ_a for some $a \in A$ always has a deterministic outcome for any event $E \subset A$, namely the event either occurs if $a \in M$ and doesn't occur if $a \notin M$. To extend these ideas, recall an arbitrary state $\rho \in D(\mathcal{H})$ is diagonalizable with respect to some orthonormal basis $\{|\psi_i\rangle\}_{i=1}^d$, such that

$$\rho = \sum_{i=1}^{d} \rho_i \,|\psi_i\rangle\langle\psi_i|\,, \tag{3.31}$$

where for all $i \in \{1, \cdots, d\}$ we have $\rho_i \geq 0$ and $\sum_{i=1}^d \rho_i = 1$. A reasonable suggestion for a quantum analogue of measuring a probability distribution $p \in \mathrm{Prob}(A)$, is to measure the quantum state ρ with respect to it's diagonalizing basis $\{|\psi_i\rangle\}_{i=1}^d$. By Born's rule we then get for all $i \in \{1, \cdots, d\}$,

$$\rho_i = \mathrm{Tr}[\rho\,|\psi_i\rangle\langle\psi_i|], \tag{3.32}$$

is the expectation of measuring the eigenvalues of the projections $|\psi_i\rangle\langle\psi_i|$. So far we have only assumed we can construct some physical experiment that allows us to measure the eigenvalues of a self adjoint operator. It turns out that since the projections $(|\psi_i\rangle\langle\psi_i|)_{i=1}^d$ are all commuting self adjoint operators, and

$$\sum_{i=1}^{d} |\psi_i\rangle\langle\psi_i| = \mathbf{1}, \tag{3.33}$$

we are also able to create a physical experiment for a basis measurement of a state ρ. In this case for all $i \in \{1, \cdots, d\}$, the eigenvalue ρ_i can be interpreted as the probability of measure the basis vector $|\psi_i\rangle$ during the basis measurement process. To return to the problem of defining a *canonical* way of measuring a state ρ, the key problem we face with using the diagonalizing basis $\{|\psi_i\rangle\}_{i=1}^d$ is that there are always other states $\rho' \in D(\mathcal{H})$ such that $\rho' \neq \rho$, and for all $i \in \{1, \cdots, d\}$,

$$\mathrm{Tr}[\rho\,|\psi_i\rangle\langle\psi_i|] = \mathrm{Tr}[\rho'\,|\psi_i\rangle\langle\psi_i|], \tag{3.34}$$

that is, from the perspective of measuring the basis $\{|\psi_i\rangle\}_{i=1}^d$ there are at least two distinct quantum states which give the same measurement outcome. We stress the point that there is no intrinsic way of measuring a quantum system. Before any measurements can be made on a quantum state one must first make a choice of which self adjoint operator $a \in B(\mathcal{H})$ or orthonormal basis $\mathcal{B}$ to measure. More generally, you must first construct a positive operator valued measure.

Definition III.3.2 (Positive Operator Valued Measure)**.** *Let Ω be a finite set. A function $\mu : \Omega \to B(\mathcal{H})$, is called a* ***positive operator valued measure*** *(POVM) if for all $\omega \in \Omega$,*

$$0 \leq \mu(\omega) \leq \boldsymbol{1}, \;\; \sum_{\omega\in\Omega} \mu(\omega) = \boldsymbol{1}. \tag{3.35}$$

A POVM generalizes the notion of creating an experiment from some self adjoint operator $a \in B(\mathcal{H})$. If $\mu : \Omega \to B(\mathcal{H})$ is a POVM then we generalize Borns rule as follows. For each POVM we assume there exists a physical experiment which takes values in Ω and

$\mathrm{Tr}[\rho\mu(\omega)]$ is the probability of measuring $\omega \in \Omega$. This formalization is sensible since for all $\omega \in \Omega$ we have $0 \leq \mu(\omega)$, which implies $\mathrm{Tr}[\rho\mu(\omega)] \geq 0$, and by linearity

$$\sum_{\omega\in\Omega} \mathrm{Tr}[\rho\mu(\omega)] = \mathrm{Tr}[\rho\mathbf{1}] = 1. \tag{3.36}$$

A key example of a POVM comes from a basis measurement set. Namely, fix any orthonormal basis $\mathcal{B} = \{|x\rangle\}_{i=1}^{d}$ of $\mathcal{H}$, and then define $\mu_{\mathcal{B}} : \{1, \cdots, d\} \to B(\mathcal{H})$ as $\mu_{\mathcal{B}}(x) = |x\rangle\langle x|$. An equivalent formalization of a basis measurement set is through a defining a pinching map $V_{\mathcal{B}} : D(\mathcal{H}) \to D(\mathcal{H})$ as

$$V_{\mathcal{B}}(\sigma) = \sum_{x=1}^{d} \lambda_x |x\rangle\langle x|, \quad \text{where } \lambda_x = \langle x|\sigma|x\rangle, \forall x \in \{1 \cdots, d\}. \tag{3.37}$$

Clearly, for any $\sigma \in D(\mathcal{H})$,

$$\mathrm{Tr}[\sigma] = \sum_{a\in A} \langle a|\sigma|a\rangle = \sum_{a\in A} \lambda_a = 1, \tag{3.38}$$

and $\langle a|\sigma|a\rangle \geq 0$ for all $x \in \{1, \cdots, d\}$, as $\sigma \geq 0$. Therefore, $V_{\mathcal{B}}(\sigma) \geq 0$, $\mathrm{Tr}[V(\sigma)] = 1$, and $V(\sigma) \in D(\mathcal{H})$. The map pinching map $V_{\mathcal{B}}$ is a CPTP map, which is also an example of a quantum-to-classical map.

Definition III.3.3 (Quantum-to-classical map)**.** *Let $\mathcal{H}, \mathcal{K}$ be two finite dimensional Hilbert spaces. A CPTP map $V : D(\mathcal{H}) \to D(\mathcal{K})$ is a* ***quantum-to-classical*** *map if there exists a basis $\mathcal{B}$ of $\mathcal{K}$ such that for all $\sigma \in D(\mathcal{H})$ the density operator $V(\sigma) \in D(\mathcal{K})$ is diagonalizable with respect to $\mathcal{B}$, that is,*

$$V(\sigma) = \sum_{|x\rangle\in\mathcal{B}} \lambda_x |x\rangle\langle x|, \quad \textit{where } \lambda_x = \langle x|V(\sigma)|x\rangle. \tag{3.39}$$

A key point to keep in mind is that every POVM $\mu : \Omega \to D(\mathcal{H})$, can be interpreted as a quantum-to-classical CPTP map. Let $k = |\Omega|$ and $\mathcal{K}$ be a k-dimensional Hilbert space while $\mathcal{B} = \{|\omega\rangle\}_{\omega\in\Omega}$ is an orthonormal basis in $\mathcal{K}$. Define a CPTP map $V_\mu : D(\mathcal{H}) \to D(\mathcal{K})$ as,

$$V_\mu(\rho) := \sum_{\omega\in\Omega} \lambda_\omega |\omega\rangle\langle\omega|, \quad \text{where } \lambda_\omega = \mathrm{Tr}[\mu(\omega)\rho]. \tag{3.40}$$

An equivalent, and perhaps more useful interpretation is that a quantum-to-classical map created by a POVM μ, is the function $V_\mu : D(\mathcal{H}) \to \mathrm{Prob}(\Omega)$ such that for all $\rho \in D(\mathcal{H})$ we define the probability distribution $V_\mu(\rho) \in \mathrm{Prob}(\Omega)$ as

$$V_\mu(\rho)(\omega) := \mathrm{Tr}[\rho\mu(\omega)]. \tag{3.41}$$

This leads us to the most important takeaway from this section: if we prepare a normal state $\rho \in D(\mathcal{H})$, and fix a measurement process (POVM) $\mu : \Omega \to D(\mathcal{H})$, then the measurement outcome, and hence measurement interaction, is described by a classical probability distribution. This showcases Kolmogorov's axioms of probability are closely related to those of quantum probability. Crucially, this does not imply quantum randomness is essentially classical randomness. There are fundamentally different things one can achieve with quantum states compared to classical probabilities.[2] But when taking measurements of a quantum system, we as observers are ultimately interacting with classical randomness.

[2] A lot of the interesting work in quantum information and computing theory relies on this fact!

III.3.2 Quantum hypothesis testing

Quantum hypothesis testing is a very active field in quantum information/probability theory. A topic of particular interest is generalizing the quantum Stein and Chernoff lemmas to more quantum hypothesis testing frameworks [17, 22, 23, 18], much like in Theorem II.2.1. Part of the success in developing the theory of quantum hypothesis testing can be attested to the similarities it shares with its classical counterpart. We begin we the following hypothesis setup. Let $\mathcal{H}$ be a finite dimensional Hilbert space and $\rho, \sigma \in D(\mathcal{H})$ be two distinct quantum states. Suppose we have a quantum state $\rho_0 \in D(\mathcal{H})$ which we know is either equal to ρ or σ. Then we have the two simple hypotheses

$$\begin{aligned} H_0 &: \rho_0 = \rho, \\ H_1 &: \rho_0 = \sigma. \end{aligned} \tag{3.42}$$

In the classical hypothesis testing, the analogous setup is deducing if a random variable X either has law p or q (2.13), where $p, q \in \text{Prob}(A)$ and A is a finite set. The first step in the classical testing process is producing a string $x \in A^n$ of observations by sampling $n \in \mathbb{N}$ i.i.d copies of X. The second step is defining a test $T_n \subset A^n$ (Definition II.2.1), which assigns the decision of accepting H_0 if $x \in T_n$ and accepting H_1 if $x \notin T_n$. We emphasis every subset $T_n \subset A^n$ defines a test, and any possible string of observations $x \in A^n$ must result in either accepting H_0 or H_1.

The quantum analogy of sampling n copies of the random variable X, is doing a measurement on the product state $\rho_0^{\otimes n} \in D(\mathcal{H}^{\otimes n})$. As we discussed in the previous section, to do any measurement on a quantum state, we are required to define a POVM. In the context of hypothesis testing the main objective is to use the observational data to deduce which of the two hypotheses to accept, therefore all suitable quantum measurements for this task are POVMs of the form $\mu : \{0, 1\} \to B(\mathcal{H}^{\otimes n})$. Since $\mu(0) + \mu(1) = \mathbf{1}^{\otimes n}$, there is only one degree of freedom in defining POVMs with two measurement outcomes namely for all finite dimensional Hilbert spaces $\mathcal{K}$,

$$\{\mu : \{0, 1\} \to B(\mathcal{K}); \mu \text{ is a POVM}\} \simeq \{T \in B(\mathcal{K}); 0 \leq T \leq \mathbf{1}\}, \tag{3.43}$$

and we can give the following definition of a quantum test.

Definition III.3.4 (Quantum Tests)**.** *Let $n \in \mathbb{N}$, a positive operator $T_n \in B(\mathcal{H}^{\otimes n})$ is called a **test** if $0 \leq T_n \leq \mathbf{1}^{\otimes n}$ and if the basis for accepting the null or alternative hypothesis* (3.42) *is the measurement outcome of the POVM $\mu : \{0, 1\} \to B(\mathcal{H}^{\otimes n})$ defined as*

$$\mu(0) = T_n, \quad \mu(1) = 1 - T_n. \tag{3.44}$$

Namely if we measure the state $\rho_0^{\otimes n}$ with respect to μ then,

$$\begin{aligned} &\text{The outcome of measuring with } \mu \text{ is } 0 \implies \text{accept } H_0, \\ &\text{The outcome of measuring with } \mu \text{ is } 1 \implies \text{Accept } H_1. \end{aligned} \tag{3.45}$$

Much like in the classical setting, a topic of central importance in studying quantum hypothesis testing are error rates. In the quantum hypothesis testing regime as detailed in (3.42), one classifies the errors for a test $T_n \in B(\mathcal{H}^{\otimes n})$ as follows, let μ_n be the POVM

associated to the test T_n as defined in Definition III.3.4. We say a false positive error occurs, or a type I error, if the outcome of the measurement of μ is 1 while $\rho_0 = \rho$, i.e. falsely accepting H_1 while H_0 is true. The probability of a type I error occurring is denoted by $\alpha(T_n)$ and given by

$$\alpha(T_n) := \mathrm{Tr}[\rho^{\otimes n}(\mathbf{1} - T_n)] = \mathrm{Tr}[\rho\mu(1)]. \tag{3.46}$$

Similarly, we say a false negative occurs, or type II error, if the measurement of the POVM μ is 0 when $\rho_0 = \sigma$. The probability of a type II error occurring is denoted by $\beta(T_n)$ and given by

$$\beta(T_n) := \mathrm{Tr}[\sigma^{\otimes n}T_n] = \mathrm{Tr}[\sigma^{\otimes n}\mu(0)]. \tag{3.47}$$

In the classical case we were interested in describing the optimal way to conclude which hypothesis is valid. In the quantum setting this endeavour is equally relevant, and we can borrow the two definitions we had for optimality of a test, namely with Chernoff's and Stein's optimality. A test $T_n \subset B(\mathcal{H}^n)$ is Chernoff optimal if it minimizes the sum of the type I and type II errors, namely,

$$\gamma(T_n) := \alpha(T_n) + \beta(T_n). \tag{3.48}$$

Recall this case treats the event a type I or type II on an equal footing. The other approach, is to say a test $0 \leq T_n \leq \mathbf{1}^{\otimes n}$ is Stein optimal if it minimizes the type II error while ensuring the type I error is bounded above by some fixed $\delta > 0$. This notion of optimally emphasis reducing the event of a type II error occurs, modelling the situation where a type II error occurring has far greater consequences than a type I error occurring. Nearly identical to the classical case, we can define the optimal errors as inspired by the Chernoff and Stein optimality, which will similarly provide exciting operational interpretations of quantum entropies. Let $n \in \mathbb{N}$ and $\delta > 0$, consider the quantum hypotheses setting described in (3.42). Let $n \in \mathbb{N}$ and with $\mathbf{1}_n$ denote the identity in $B(\mathcal{H}^{\otimes n})$. Define

$$\beta^*_{N,\varepsilon} := \inf\{\beta(T); T \in B(\mathcal{H}^{\otimes n}), 0 \leq T \leq \mathbf{1}_n \text{ given } \alpha(T) < \varepsilon\}, \tag{3.49}$$

which is the type II error for the Stein optimal test. Similarly, define

$$\gamma_n{}^* := \inf\{\alpha(T) + \beta(T); T \in B(\mathcal{H}^{\otimes n}), 0 \leq T \leq \mathbf{1}_n\}, \tag{3.50}$$

which is the total error for the Chernoff optimal test. We can now state both the quantum Stein lemma and the quantum Chernoff lemma.

Lemma III.3.1 (Quantum Stein's Lemma)**.** *Let $\mathcal{H}$ be a finite dimensional Hilbert space and $\varepsilon \in (0,1)$. Suppose $\rho, \sigma \in D(\mathcal{H})$ are normal states, and consider the definition of $\beta^*_{n,\varepsilon}$ in* (3.49)*. Then*

$$\lim_{n\to\infty} \frac{1}{n} \log \beta^*_{n,\varepsilon} = -D(\rho||\sigma). \tag{3.51}$$

Moreover, define the **quantum Chernoff relative entropy** $C : D(\mathcal{H}) \times D(\mathcal{H}) \to [0, \infty)$ as

$$C_q(\rho, \sigma) := \sup_{\alpha\in(0,1)} (1 - \alpha)R_\alpha(\rho||\sigma), \tag{3.52}$$

where R_α is the Petz-Rényi relative entropy defined in Definition III.1.9.

Lemma III.3.2 (Quantum Chernoff's Lemma)**.** *Let $\mathcal{H}$ be a finite dimensional Hilbert space. Suppose $\rho, \sigma \in D(\mathcal{H})$ are normal states, and consider the definition of γ_n^* in* (3.50)*. Then*

$$\lim_{n\to\infty} \frac{1}{n} \log \gamma_n^* = -C_q(\rho, \sigma). \tag{3.53}$$

Proof. See [4]. □

In the next section we will provide a complete proof of the quantum Stein lemma. For now we just wish to admire the similarities these results share with their classical counterparts, which goes beyond the obvious structural similarities. If we take two commuting normal states $\sigma, \rho \in D(\mathcal{H})$, then as discussed in Section III.2, ρ, σ are equivalent to two probability distributions p_ρ and q_σ. In fact,

$$D_{KL}(p_\rho || p_\sigma) = D(\rho||\sigma), \text{ and } C(p_\rho, p_\sigma) = C_q(\rho, \sigma), \tag{3.54}$$

and interestingly, we can conclude: quantum tests are not more efficient at separating classical distributions than classical tests are at separating classical distributions. Additionally, consider a POVM $\mu : \Omega \to D(\mathcal{K})$, and it's quantum-to-classical map $V_\mu : D(\mathcal{H}) \to D(\mathcal{K})$ described in (3.40). Then for two non-commuting normal state $\rho, \sigma \in D(\mathcal{H})$, by the monotonicity of the Umegaki relative entropy Proposition III.1.3, we have

$$D_{KL}(V_\mu(\rho)||V_\mu(\sigma)) < D(\rho||\sigma). \tag{3.55}$$

And hence, reducing two non-commuting quantum states to classical probabilities always make them harder to distinct. However, by the quantum Stein's lemma we know the following. Let $\varepsilon > 0$, by definition of the infimum for all $n \in \mathbb{N}$ there exists a test T_n, satisfying $0 \leq T_n \leq \mathbf{1}_n$, and $\alpha(T_n) < \varepsilon$ such that

$$\beta(T_n) < \beta^*_{N,\varepsilon}\left(1 + \frac{1}{n}\right) \implies \mathrm{Tr}[\sigma^{\otimes n} T_n] \leq \left(1 + \frac{1}{n}\right) \beta^*_{\varepsilon,N}. \tag{3.56}$$

Take the sequence of tests $(T_n)_{n\in\mathbb{N}}$ satisfying (3.56) then, since using that log

$$\lim_{n\to\infty} \frac{1}{n} \log \left(\left(1 + \frac{1}{n}\right) \beta^*_{\varepsilon,n} \right) = \lim_{n\to\infty} \log \beta^*_{\varepsilon,n}, \tag{3.57}$$

from (3.56) we obtain,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}\left[\sigma^{\otimes n} T_n\right] \leq \limsup_{n\to\infty} \frac{1}{n} \log \beta^*_{\varepsilon,n} = -D(\rho||\sigma). \tag{3.58}$$

On the other hand since $\beta^*_{\varepsilon,N} \leq \beta(T_n)$ for all $n \in \mathbb{N}$ we have

$$-D(\rho||\sigma) = \liminf_{n\to\infty} \frac{1}{n} \log \beta^*_{\varepsilon,n} \leq \lim \liminf_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}\left[\sigma^{\otimes n} T_n\right], \tag{3.59}$$

and combining both results we see there exists a sequence of tests $(T_n)_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}\left[\sigma^{\otimes n} T_n\right] = -D(\rho||\sigma). \tag{3.60}$$

We can then observe that for all $\delta > 0$, there exists an $M \in \mathbb{N}$ such that for all $n \geq M$, there is a test $T_n \in B(\mathcal{H}^{\otimes n})$ such that

$$\exp\left(-n(D(\rho||\sigma)+\delta\right) \leq \mathrm{Tr}\left[\sigma^{\otimes n}T_n\right] \leq \exp\left(-n(D(\rho||\sigma)-\delta\right). \tag{3.61}$$

If we return to the (3.42) setup, for $n \geq M$ the tests T_n insure that if the underlying state $\rho_0 = \sigma$, then the probability the measurement outcome results in one falsely concluding the underlying state is ρ is bounded by

$$\exp\left(-n(D(\rho||\sigma)+\delta\right) \leq \beta(T_n) \leq \exp\left(-n(D(\rho||\sigma)-\delta\right). \tag{3.62}$$

And we arrive at a crucial difference between hypothesis testing in the classical versus quantum setting. In the classical case, sampling the product measure $q^{\otimes n}$ is equivalent to sampling n copies of a random variable with law q. However, in the quantum case there are tests $0 \leq T_n \leq \mathbf{1}_N$, such that measuring $\sigma^{\otimes n}$ with T_n is not reproducible by measuring n copies of σ.

III.3.3 Proof of the quantum Stein lemma

In this section we prove the Quantum Stein Lemma III.3.1. We follow the work of [5], which provides a more streamline approach than the original approach in [20, 32]. We begin by stating two necessary lemmas. Let $\mathcal{H}$ be a d-dimensional Hilbert space. For all normal states $\rho, \sigma \in D(\mathcal{H})$ with $\ker(\sigma) \subseteq \ker(\rho)$ define the the **von Neumann entropy** $S : D(\mathcal{H}) \to [0, \infty)$ as

$$S(\rho) := -\mathrm{Tr}[\rho \log \rho], \tag{3.63}$$

and the **quantum cross entropy** $\Delta : D(\mathcal{H}) \times D(\mathcal{H}) \to [0, \infty)$,

$$\Delta(\rho||\sigma) := S(\rho) + D(\rho||\sigma) = -\mathrm{Tr}[\rho \log \sigma]. \tag{3.64}$$

Notice $\Delta(\rho||\rho) = S(\rho)$. Consider the spectral decomposition of σ,

$$\sigma = \sum_{i=1}^{d} \lambda_i \left|i\right\rangle \left\langle i\right|. \tag{3.65}$$

Define $E_i := \left|i\right\rangle \left\langle i\right|$ for $i \in \{1, \cdots, d\}$, as the one dimensional projections onto the eigenvectors of σ. We know that $E_j E_i = E_j E_i = \delta_{ij}$ for $i, j \in \{1, \cdots, d\}$.
Let $[d] := \{1, \cdots, d\}$ and $[d]^n$ be the n-fold Cartesian product of $[d]$. For $x = (x_1, \cdots, x_n) \in [d]^n$, define $\lambda_x \in \mathbb{R}$ and $E_x \in B(\mathcal{H}^{\otimes n})$ as:

$$\lambda_x := \prod_{i=1}^{n} \lambda_{x_i}, \qquad E_x := \bigotimes_{i=1}^{n} E_{x_i}. \tag{3.66}$$

We emphasis for all $x \in [d]^n$, $\lambda_x = \lambda_x(\sigma)$, and $E_x = E_x(\sigma)$, are both dependent on σ. Additionally, for $\delta > 0$ define $T_{\delta,n}(\rho, \sigma) \subseteq [d]^n$ as follows,

$$T_{\delta,n}(\rho,\sigma) := \Big\{ x \in [d]^n; \lambda_x > 0, \left|\Delta(\rho||\sigma) + \frac{1}{n}\log\lambda_x\right| < \delta \Big\}, \quad \text{where } \lambda_x = \lambda_x(\sigma). \tag{3.67}$$

Finally, we construct the projection $P_{\delta,n}(\rho, \sigma) \in B(\mathcal{H})$,

$$P_{\delta,n}(\rho,\sigma) := \sum_{x \in T_{\delta,n}(\rho,\sigma)} E_x. \tag{3.68}$$

We can confirm $P_{\delta,n}(\rho,\sigma)$ is a projection by noticing for any two $x, y \in [d]^n$,

$$E_x E_y = E_y E_x = \prod_{i=1}^{n} \delta_{x_i,y_i} = \delta_{x,y}, \tag{3.69}$$

and $(P_{\delta,n}(\rho,\sigma))^2 = P_{\delta,n}$, while $P_{\delta,n}(\rho,\sigma)$ is clearly self-adjoint. We can now state and prove the most essential lemma needed in the proof of the quantum Stein Lemma.

Lemma III.3.3. *For all $n \in \mathbb{N}$, $\rho, \sigma \in D(\mathcal{H})$ with $\ker(\sigma) \subseteq \ker(\rho)$, and $\delta > 0$ we have:*

1. $P_{\delta,n}(\rho,\sigma)\sigma^{\otimes n} = \sigma^{\otimes n} P_{\delta,n}(\rho,\sigma)$,
2. $P_{\delta,n}(\rho,\sigma)\sigma^{\otimes n} P_{\delta,n}(\rho,\sigma) \leq e^{-n(\Delta(\rho||,\sigma)-\delta)} P_{\delta,n}(\rho,\sigma)$,
3. $P_{\delta,n}(\rho,\sigma)\sigma^{\otimes n} P_{\delta,n}(\rho,\sigma) \geq e^{-n(\Delta(\rho||\sigma)+\delta)} P_{\delta,n}(\rho,\sigma)$,
4. $\lim_{n\to\infty} \mathrm{Tr}[\rho^{\otimes n} P_{\delta,n}(\rho,\sigma)] = 1$.

Proof. 1. Recall the spectral decomposition of σ in (3.65), and for $x \in [d]^n$, recall the projection $E_x = E_x(\sigma)$ constructed in (3.66). We can then write $\sigma^{\otimes n}$ as:

$$\begin{aligned}\sigma^{\otimes n} &= \sum_{y\in[d]^n} \lambda_{y_1} E_{y_1} \otimes \cdots \otimes \lambda_{y_n} E_{y_n} \\ &= \sum_{y\in[d]^n} \lambda_y E_y. \end{aligned} \tag{3.70}$$

As stated in (3.69), for all $x, y \in [d]^n$, $E_x E_y = E_y E_x = \delta_{x,y}$ and

$$\begin{aligned} P_{\delta,n}(\rho,\sigma)\sigma^{\otimes n} &= \sum_{y\in[d]^n} \sum_{x\in T_{\delta,n}(\rho,\sigma)} \lambda_y E_x E_y \\ &= \sum_{y\in[d]^n} \sum_{x\in T_{\delta,n}(\rho,\sigma)} \lambda_y E_y E_x \\ &= \sum_{x\in T_{\delta,n}(\rho,\sigma)} \lambda_x E_x = \sigma^{\otimes n} P_{\delta,n}(\rho,\sigma), \end{aligned} \tag{3.71}$$

hence $P_{\delta,n}(\rho,\sigma)$ and $\sigma^{\otimes n}$ commute.
2 and *3*. We begin with noticing for $x \in T_{\delta,n}(\rho,n)$,

$$|\Delta(\rho||\sigma) + \frac{1}{n}\log\lambda_x| < \delta \iff e^{-n(\Delta(\rho||\sigma)+\delta)} < \lambda_x < e^{-n(\Delta(\rho||\sigma)-\delta)}. \tag{3.72}$$

By (3.71) and *1*, we obtain

$$P_{\delta,n}(\rho,\sigma)\sigma^{\otimes n} P_{\delta,n}(\rho,\sigma) = P_{\delta,n}(\rho,\sigma)\sigma^{\otimes n} = \sum_{x\in T_{\delta,n}(\rho,\sigma)} \lambda_x E_x. \tag{3.73}$$

Hence,

$$\begin{aligned} P_{\delta,n}(\rho,\sigma)\sigma^{\otimes n} P_{\delta,n}(\rho,\sigma) &= \sum_{x\in T_{\delta,n}(\rho,\sigma)} \lambda_x E_x \\ &\leq \sum_{x\in T_{\delta,n}(\rho,\sigma)} e^{-n(\Delta(\rho||\sigma)-\delta)} E_x \\ &= e^{-n(\Delta(\rho||\sigma)-\delta)} P_{\delta,n}(\rho,\sigma). \end{aligned} \tag{3.74}$$

An identical argument leads us to concluding

$$P_{\delta,n}(\rho,\sigma)\sigma^{\otimes n}P_{\delta,n}(\rho,\sigma) \geq e^{-n(\Delta(\rho||\sigma)+\delta)}P_{\delta,n}(\rho,\sigma). \tag{3.75}$$

4. We will make use of the weak law of large numbers (Theorem Theorem A.2.5). Let X be the random variable on $[d]$ with law $p_\rho \in \text{Prob}([d])$, which is defined as

$$p_\rho(i) := \text{Tr}[\rho E_i], \qquad\qquad i \in [d]. \tag{3.76}$$

We can confirm that p_ρ is a probability distribution since ρ is a density operator and $\sum_{i=1}^d E_i = \mathbf{1}$. Consider the i.i.d set of random variables $(X_i)_{i\in\mathbb{N}}$ each with law p_ρ. Then for $x \in [d]^n$,

$$p_\rho^n(X_1 = x_1, \cdots, X_n = x_n) = \text{Tr}[\rho^{\otimes n}E_x] = \prod_{i=1}^n \text{Tr}[\rho E_{x_i}]. \tag{3.77}$$

For $i \in \{1, \cdots, n\}$, define the i.i.d random variables U_i on $\mathbb{R}$ as,

$$U_i := -\log(\lambda_{X_i}). \tag{3.78}$$

Then we have

$$\begin{aligned}\mathbb{E}[U_i] &= -\sum_{i=1}^d p_\rho(U_i = \log\lambda_{x_i})\log\lambda_{x_i} = -\sum_{i=1}^d p_\rho(X_i = x_i)\log\lambda_{x_i} \\ &= -\sum_{i=1}^d \text{Tr}[\rho E_{x_i}]\log\lambda_{x_i} = -\text{Tr}[\rho\log(\sigma)] = \Delta(\rho||\sigma).\end{aligned} \tag{3.79}$$

Where we use linearity of the trace and functional calculus of positive operators. Now observe,

$$p_\rho^n\Big(\Big|\frac{1}{n}\sum_{i=1}^n U_i - \Delta(\rho||\sigma)\Big| < \delta\Big) = p_\rho^n\Big((X_1, \cdots, X_n) \in T_{\delta,n}(\rho,\sigma)\Big), \tag{3.80}$$

because for all $x = (x_1, \cdots, x_n) \in [d]^n$,

$$\Big|\Delta(\rho||\sigma) + \frac{1}{n}\log(\lambda_x)\Big| = \Big|\frac{1}{n}\sum_{i=1}^n \log(\lambda_{x_i}) + \Delta(\rho||\sigma)\Big| = \Big|-\frac{1}{n}\sum_{i=1}^n \log(\lambda_{x_i}) - \Delta(\rho||\sigma)\Big|. \tag{3.81}$$

Therefore,

$$\begin{aligned}p_\rho^n\Big(\Big|\frac{1}{n}\sum_{i=1}^n U_i - \Delta(\rho||\sigma)\Big| < \delta\Big) &= \sum_{x\in T_{\delta,n}(\rho,\sigma)} p_\rho(X_1 = x_1, \cdots, X_n = x_n) \\ &= \sum_{x\in T_{\delta,n}(\rho,\sigma)} \text{Tr}[\rho^{\otimes n}E_x] = \text{Tr}[\rho^{\otimes n}P_{\delta,n}(\rho,\sigma)].\end{aligned} \tag{3.82}$$

But then by the weak law of large numbers we have for all $\delta > 0$,

$$\lim_{n\to\infty} \text{Tr}[\rho^{\otimes n}P_{\delta,n}(\rho,\sigma)] = 1. \tag{3.83}$$

□

Lemma III.3.4. *Let $\mathcal{K}$ be a finite dimensional Hilbert space. Assume $P, Q \in B(\mathcal{K})$ are tests and $\tau \in D(\mathcal{K})$. Then*

1. $$\mathrm{Tr}[\tau PQP] \geq \mathrm{Tr}[\tau Q] - 2\sqrt{\mathrm{Tr}[\tau(\mathbf{1} - P)]}. \tag{3.84}$$

2. *Moreover, if $U \in B(\mathcal{K})$ is a projection with the property that $\tau = U\tau$ and $(c\mathbf{1} - \tau)U \geq 0$ for some $c > 0$, then*

$$\mathrm{Tr}[PQP] \geq \frac{1}{c}\Big(\mathrm{Tr}[\tau P] - 2\sqrt{\mathrm{Tr}[\tau(\mathbf{1} - P)]} - \mathrm{Tr}[\tau(\mathbf{1} - U)]\Big). \tag{3.85}$$

Proof. See [5, Appendix A]. □

Proof of Quantum Stein's Lemma III.3.1. Step 1: We will first show that

$$\limsup_{n\to\infty} \frac{1}{n}\beta_{\varepsilon,n} \leq -D(\rho||\sigma). \tag{3.86}$$

Let $\delta > 0$, since $\ker(\sigma) \subseteq \ker(\rho)$ we can apply Lemma III.3.3. To simplify notation we let let $P_{\delta,n}(\rho,\rho) = P_{\delta,n}(\rho)$ for all $n \in \mathbb{N}$. Moving on by Lemma III.3.4 we have,

$$\mathrm{Tr}[\rho^{\otimes} nP_{\delta,n}(\rho,\sigma)P_{\delta,n}(\rho)P_{\delta,n}(\rho,\sigma)] \geq Tr[\rho^{\otimes n}P_{\delta,n}(\rho)] - 2\sqrt{\mathrm{Tr}[\rho^{\otimes n}(\mathbf{1} - P_{\delta,n}(\rho,\sigma))]}. \tag{3.87}$$

Then by part *4* of Lemma III.3.1, there exists an $M \in \mathbb{N}$ such that for all $n \geq M$,

$$\Big|\mathrm{Tr}[\rho^{\otimes n}P_{\delta,n}(\rho,\sigma)] - 1\Big| = \Big|\mathrm{Tr}[\rho^{\otimes n} - \rho^{\otimes n}P_{\delta,n}(\rho,\sigma)]\Big| < \frac{\varepsilon}{2}. \tag{3.88}$$

Clearly $\mathrm{Tr}[\rho^{\otimes n}P_{\delta,n}(\rho)] \leq 1$ as $P_{\delta,n}(\rho)$ is a projection and by (3.87),

$$\mathrm{Tr}[\rho^{\otimes n}P_{\delta,n}(\rho,\sigma)P_{\delta,n}(\rho)P_{\delta,n}(\rho,\sigma)] \geq 1 - \epsilon. \tag{3.89}$$

Applying part *1* of Lemma III.3.3,

$$\sigma^{\otimes n}P_{\delta,n}(\rho,\sigma) = P_{\delta,n}(\rho,\sigma)\sigma^{\otimes n} = P_{\delta,n}(\rho,\sigma)\sigma^{\otimes n}P_{\delta,n}(\rho,\sigma). \tag{3.90}$$

By part 2 of Lemma III.3.3,

$$\sigma^{\otimes n}P_{\delta,n}(\rho,\sigma) \leq e^{-n(\Delta(\rho||\sigma)-\delta)}P_{\delta,n}(\rho,\sigma). \tag{3.91}$$

Furthermore, in general if $A \leq B$ for two operators then $PAP \leq PBP$ for a projection P. Hence,

$$P_{\delta,n}(\rho)\sigma^{\otimes n}P_{\delta,n}(\rho,\sigma)P_{\delta,n}(\rho) \leq e^{-n(\Delta(\rho||\sigma)-\delta)}P_{\delta,n}(\rho)P_{\delta,n}(\rho,\sigma)P_{\delta,n}(\rho). \tag{3.92}$$

Taking the trace of (3.92), and applying the cyclic property we find

$$\begin{aligned}\mathrm{Tr}[\sigma^{\otimes n}P_{\delta,n}(\rho,\sigma)P_{\delta,n}(\rho)P_{\delta,n}(\rho,\sigma)] =& \mathrm{Tr}[P_{\delta,n}(\rho)\sigma^{\otimes n}P_{\delta,n}(\rho,\sigma)P_{\delta,n}(\rho)] \\ \leq& e^{-n(\Delta(\rho||\sigma)-\delta)}\mathrm{Tr}[P_{\delta,n}(\rho,\sigma)P_{\delta,n}(\rho)] \\ \leq& e^{-n(\Delta(\rho||\sigma)-\delta)}\mathrm{Tr}[P_{\delta,n}(\rho)].\end{aligned} \tag{3.93}$$

Using part *3* of Lemma III.3.1, we find

$$P_{\delta,n}(\rho) \leq e^{n(S(\rho)+\delta)} P_{\delta,n}(\rho)\rho^{\otimes n} P_{\delta,n}(\rho), \tag{3.94}$$

where we use that $\Delta(\rho||\rho) = S(\rho)$. Taking the trace (3.94) we get

$$\begin{aligned} \mathrm{Tr}[P_{\delta,n}(\rho)] &\leq e^{n(S(\rho)+\delta)} \mathrm{Tr}[P_{\delta,n}(\rho)\rho^{\otimes n} P_{\delta,n}(\rho)] \\ &\leq e^{n(S(\rho)+\delta)}. \end{aligned} \tag{3.95}$$

Then, combining (3.93) and (3.95) we find

$$\mathrm{Tr}[\sigma^{\otimes n} P_{\delta,n}(\rho,\sigma) P_{\delta,n}(\rho) P_{\delta,n}(\rho,\sigma)] \leq e^{n(S(\rho)+\delta)} e^{-n(\Delta(\rho||\sigma)-\delta)} = e^{-n(D(\rho||\sigma)-2\delta)}. \tag{3.96}$$

Where we use that $S(\rho) - \Delta(\rho||\sigma) = -D(\rho||\sigma)$. Now we can define the sequence of tests

$$T_n := P_{\delta,n}(\rho,\sigma) P_{\delta,n}(\rho) P_{\delta,n}(\rho,\sigma), \tag{3.97}$$

for all $n \in \mathbb{N}$. By equation (3.89), we have that T_n satisfies, $\mathrm{Tr}[\rho^{\otimes n} T_n] > 1 - \epsilon$ for all $n \geq M$, hence $\alpha(T_n) < \varepsilon$. Therefore,

$$\limsup_{n\to\infty} \frac{1}{n} \log \beta_{\varepsilon,n} \leq \limsup_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} T_n] \leq -D(\rho||\sigma) + 2\delta. \tag{3.98}$$

Since $\delta > 0$ was arbitrary,

$$\limsup_{n\to\infty} \frac{1}{n} \log \beta_{\varepsilon,n} \leq -D(\rho||\sigma). \tag{3.99}$$

Step 2: What is left to prove is that

$$\liminf_{n\to\infty} \frac{1}{n} \beta_{\varepsilon,n} \geq -D(\rho||\sigma). \tag{3.100}$$

The proof of this claim is remarkably similar to what we have done in *Step 1*, and we refer the reader to [5] for the rest of the argument. Combining both of the steps it follows that

$$\lim_{n\to\infty} \frac{1}{n} \beta_{\varepsilon,n} = -D(\rho||\sigma). \tag{3.101}$$

□

III.4 The first quantum Sanov theorem

In this section we turn our focus to the central topic of this thesis, which is, understanding and producing a quantum version of Sanov's theorem. The first step in developing this theory was taken in 2005 when Bjelaković et al. published *A quantum version of Sanov's Theorem* [6]. This paper is highly influential and shapes much of our own research. However, we refrain from calling the result a *proper* extension of the Sanov theorem. Rather, we refer to it as the Operational Quantum Sanov Theorem. We now state the major result of [6].

Theorem III.4.1 (Operational quantum Sanov Theorem)**.** *For fixed a state $\sigma \in D(\mathcal{H})$ and a open collection of states $\mathcal{A} \subset D(\mathcal{H})$, there exists a sequence $\{P_n\}_{n\in\mathbb{N}}$ of projections $P_n \in B(\mathcal{H}^{\otimes n})$ such that*

$$\lim_{n\to\infty} \mathrm{Tr}[\rho^{\otimes n} P_n] = 1, \quad \forall \rho \in \mathcal{A} \tag{3.102}$$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} P_n] = -\inf_{\rho\in\mathcal{A}} D(\rho||\sigma). \tag{3.103}$$

Moreover, for all sequences of projections $\{Q_n\}_{n\in\mathbb{N}}$ satisfying $\lim_{n\to\infty} \psi^{\otimes n}(Q_n) = 1$ for all $\rho \in \mathcal{A}$, we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}\left[\sigma^{\otimes n} Q_n\right] \geq -\inf_{\rho\in\mathcal{A}} D(\rho||\sigma). \tag{3.104}$$

There are several features about this result that makes it appear like a valid extension of the classical Sanov theorem. For one, the classical setting the Kullback-Leibler relative entropy appears both in Sanov's theorem and Stein's lemma. Likewise, the Umegaki relative entropy appears in the quantum Stein lemma and in Theorem III.4.1. Moreover, for any open sequence of normal states $\mathcal{A} \subseteq D(\mathcal{H})$, if $\rho \in \mathcal{A}$ then the sequence of tests P_n constructed in Theorem III.4.1 ensuring

$$\lim_{n\to\infty} \mathrm{Tr}[\rho^{\otimes n} P_n] = 1. \tag{3.105}$$

Which is very similar to the classical result Proposition II.2.2, that states for any open $\Omega \subset \mathrm{Prob}(A)$, if $q \in \Omega$, then

$$\lim_{n\to\infty} q^n(L_n \in \Omega) = 1, \tag{3.106}$$

where L_n is the empirical measure. This in a sense suggests that P_n behaves very similarly to classical test $L_n^{-1}(\Omega)$, however, there is one crucial difference. As we will discuss in much greater detail in Chapter IV, the sequence of tests $P_n := P_n(\mathcal{A}, \sigma)$ is dependent on the set $\mathcal{A}$, and quantum state $\sigma \in D(\mathcal{H})$. In contrast the set of classical tests $T_n := L_n^{-1}(\Omega)$ only depends on the collection $\Omega \subset \mathrm{Prob}(A)$, and for all $q \in \mathrm{Prob}(X)$,

$$\lim_{n\to\infty} \frac{1}{n} \log q^n(L_n^{-1}(\Omega)) = -\inf_{p\in\Omega} D_{KL}(p||q). \tag{3.107}$$

If we compare it to (3.60), which is a consequence of the quantum Stein lemma, then we see Theorem III.4.1 is in fact a quantum extension of the operational Sanov Theorem II.2.1.

Sketch of Proof of Theorem III.4.1. The proof can be split into two steps. In the first step we prove that for any sequence of projections $Q_{n\in\mathbb{N}}$, satisfying $\lim_{n\to\infty} \mathrm{Tr}[\rho^{\otimes n} Q_n] = 1$ for all $\rho \in \mathcal{A}$, we have

$$-\inf_{\rho\in\mathcal{A}} D(\rho||\sigma) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} Q_n]. \tag{3.108}$$

In the second step, a sequence of tests $(P_n)_{n\in\mathbb{N}}$ is constructed such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} P_n] \leq -\inf_{\rho\in\mathcal{A}} D(\rho||\sigma). \tag{3.109}$$

Applying (3.108) to the sequence of projections $(P_n)_{n\in\mathbb{N}}$ then finishes the proof.
Step 1: Due to the quantum Stein's lemma (Lemma III.3.1) we have that for any sequence

of projections $\{Q_n\}_{n\in\mathbb{N}}$ satisfying $Q_n \in B(\mathcal{H})^{\otimes n}$, and $\lim_{n\to\infty} \mathrm{Tr}[\rho^{\otimes n}(Q_n)] = 1$ for all $\rho \in \mathcal{A}$, it holds that

$$-D(\rho||\sigma) \leq -\inf_{\rho\in\mathcal{A}} D(\rho||\sigma) \leq \lim_{n\to\infty} \frac{1}{n}\beta_{\epsilon,n} \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} Q_n]. \tag{3.110}$$

Note, for large enough n the sequence Q_n would satisfy $\rho^{\otimes n}(Q_n) \geq 1-\epsilon$. Since this holds for all $\rho \in \mathcal{A}$ we clearly have that

$$-\inf_{\rho\in\mathcal{A}} D(\rho||\sigma) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} Q_n]. \tag{3.111}$$

Step 2: Just as in the classical case (Theorem II.2.1, the second step requires quite a lot more effort to prove. We refer the reader to [6] for the argument. □

III.5 Another quantum Sanov theorem

Here we provide an account of [16], which offers an alternative approach to developing a quantum Sanov Theorem. We will refer to the proposed quantum Sanov Theorem in [16], as Hayashi's quantum Sanov Theorem. To state the result of Hayashi we will require some representation theory. Therefore, in the appendix (Section A.3) we provide a brief overview of the relevant theory. A key tool we will use is Schur-Weyl duality, which is closely related to the following two groups: S_n, the group of n-permutations, and $U(d)$, the set of $d \times d$ complex unitary matrices. We will consider the following two representations on each of these two groups.

Definition III.5.1. *Let $n, d \in \mathbb{N}$. Define the map $P : S_n \to (\mathbb{C}^{d\otimes d})^{\otimes n}$ as follows. For $|i_1\rangle, \cdots, |i_n\rangle \in \mathbb{C}^d$, and $\pi \in S_n$ define*

$$P(\pi)(|i_1\rangle \otimes \cdots \otimes |i_n\rangle) := \left|i_{\pi^{-1}(1)}\right\rangle \otimes \cdots \otimes \left|i_{\pi^{-1}(n)}\right\rangle. \tag{3.112}$$

We will also make use of the tensor product representation of the unitary group $U(d)$ for $d \in \mathbb{N}$.

Definition III.5.2. *Define the map $R : U(d) \to \left(\mathbb{C}^{d\times d}\right)^{\otimes n}$ as $R(U) := U^{\otimes n}$.*

It is well know, for example as shown in [14], that P is a representation of S_n and R is a representation of the group $U(d)$. We will make use of the fact representations P and R can be reduced into irreducible representations (irreps) which can be labelled by young diagrams $\lambda \vdash_d n$ (Definition A.3.2). For $n, d \in \mathbb{N}$, let $Y_{n,d}$ be the set of all Young diagrams satisfying $\lambda \vdash_d n$. We will denote the set of irreps decomposing P with $P_\lambda : S_n \to \mathrm{GL}(V_\lambda)$, and the set of irreps of decomposing R by $R_\lambda : U(d) \to \mathrm{GL}(W_\lambda)$. By Schur-Weyl duality (Theorem A.3.3) we know

$$\mathcal{H}^{\otimes n} \simeq \bigoplus_{\lambda\in Y_{n,d}} V_\lambda \otimes W_\lambda. \tag{3.113}$$

Before we make use of this decomposition of $\mathcal{H}^{\otimes n}$ there are several definitions we must provide.

Fix an orthonormal basis $\mathcal{B} = \{|i\rangle \langle i|\}_{i=1}^{d}$ of $\mathcal{H}$. For $n \in \mathbb{N}$, first define the set of normal states $\rho \in D(\mathcal{H})$ which are diagonalizable by $\mathcal{B}$ as

$$D[\mathcal{B}] := \left\{ \rho \in D(\rho); \rho = \sum_{i=1}^{d} \lambda_i |i\rangle \langle i| \right\}. \tag{3.114}$$

Additionally, define the **empirical states** $D_n[\mathcal{B}] \subset D(\mathcal{H})$ as

$$D_n[\mathcal{B}] := \left\{ \sum_{i=1}^{d} \frac{m_i}{n} |i\rangle \langle i| \, ; \sum_{i=1}^{n} m_i = n, m_i \geq 0 \right\}, \tag{3.115}$$

and define the (basis measurement) POVM $\mu : [d] \to B(\mathcal{H})$ as

$$\mu(i) := |i\rangle \langle i| \, , \qquad \forall i \in [d]. \tag{3.116}$$

Fix a state $\rho \in D(\mathcal{H})$, then upon preforming a measurement with μ, one will obtain a value $x_1 \in [d]$. If one repeatedly measures n copies of ρ with μ, then the outcome of observations is a string $x = (x_1, \cdots, x_n) \in [d]^n$, analogous to the classical case. For each string $x = (x_1, \cdots, x_n) \in [d]^n$, define the function $\bar{\rho}_n : [d]^n \to D(\mathcal{H})$,

$$\bar{\rho}(x) := \sum_{k=1}^{n} \frac{1}{n} |i_k\rangle \langle i_k| \, . \tag{3.117}$$

Clearly for all $n \in \mathbb{N}$, and $x \in [d]^n$, $\bar{\rho}(x) \in D_n[\mathcal{B}]$. For $\rho \in D(\mathcal{H})$ define the projection $Q(\rho)$ on $\mathcal{H}^{\otimes n}$ as,

$$\begin{aligned} Q(\rho) &:= \sum_{x \in [d]^n ; \bar{\rho}_n(x) = \rho} |i_1\rangle \langle i_1| \otimes \cdots \otimes |i_n\rangle \langle i_n| \, , \qquad \rho \in D_n[\mathcal{B}], \\ Q(\rho) &:= 0, \qquad \text{otherwise.} \end{aligned} \tag{3.118}$$

Now returning to the decomposition of $\mathcal{H}^{\otimes n}$ in (3.113), let $Q_\lambda \in B(\mathcal{H}^{\otimes n})$ be the projection onto the subspace $V_\lambda \otimes W_\lambda$, and define the test,

$$T(\lambda, \rho) := Q(\lambda) Q(\rho). \tag{3.119}$$

On the set $\mathrm{Prob}([d])$ we have the following important partial ordering $\prec$:

Definition III.5.3. *For $p, q \in \mathrm{Prob}([d])$ we write*

$$p \prec q \iff \sum_{i=1}^{k} p(i) \leq \sum_{i=1}^{k} q(i), \forall k \in [d]. \tag{3.120}$$

For a normal state $\rho \in D(\mathcal{H})$, let $\lambda_1 \leq \cdots \leq \lambda_d$ be its eigen values. Define the probability distribution $\mathrm{pr}(\rho) \in \mathrm{Prob}([d])$ as

$$p(\rho)(i) := \lambda_i, \forall i \in [d]. \tag{3.121}$$

Let $\mathcal{P} \subseteq \mathrm{Prob}([d])$ be the set of probability distributions $p \in \mathrm{Prob}([d])$ such that $p(1) \leq \cdots \leq p(d)$, and let $\mathcal{P}_n := \mathcal{P} \cap \mathrm{Prob}_n([d])$, where $\mathrm{Prob}_n([d])$ is defined in (2.40).

Notice for $p \in \mathcal{P}_n$, we can define a young diagram $\lambda_p \in Y_{n,d}$ as $\lambda_p := (np(1), \cdots, np(d))$. Now define the set

$$R_n[\mathcal{B}] := \{(p, \rho) \in \mathcal{P}_n \times D_n[\mathcal{B}]; \mathrm{pr}(\rho) \prec p\}. \tag{3.122}$$

And for $r > 0$ and $\rho \in D[\mathcal{B}]$, define the set $\mathcal{C}_{r,\sigma} \subseteq R_n[\mathcal{B}]$

$$\mathcal{C}_{r,\rho} := \{(p, \rho') \in R_n[\mathcal{B}]; D(\rho'||\rho) + S(\rho') - S(p) \leq r\}. \tag{3.123}$$

The last ingredient we need for stating Hayashi's Quantum Sanov Theorem is a new type of relative entropy. For $s \in [0, \infty)\backslash\{1\}$, define the **sandwiched Rényi relative entropy** $\phi_s : D(\mathcal{H}) \times D(\mathcal{H}) \to [0, \infty]$ as

$$\phi_s(\rho||\sigma) := \begin{cases} \log \mathrm{Tr}[\sigma^{\frac{s}{2(s-1)}} \rho \sigma^{\frac{s}{2(1-s)}}]^{1-s}. & \text{if} \ker(\sigma) \subseteq \ker(\rho) \\ \infty, & \text{otherwise.} \end{cases} \tag{3.124}$$

For $r \in (0, \infty)$, define **Hayashi's relative entropy** $H_r : D(\mathcal{H}) \times D(\mathcal{H}) \to [0, \infty]$ as

$$H_r(\rho||\sigma) := \sup_{s \in (0,1)} \frac{-(1-s)r - \phi_{1-s}(\rho||\sigma)}{s}. \tag{3.125}$$

Theorem III.5.1 (Hayashi's Quantum Sanov Theorem)**.** *Fix a basis $\mathcal{B}$ and let $\rho \in D[\mathcal{B}]$. Let $r > 0$ and let $\mathcal{C}_r$ be constructed as in* (3.123)*. Then for every state $\sigma \in D(\mathcal{H})$,*

$$\lim_{n\to\infty} \mathrm{Tr}[\rho^{\otimes n} Q_n] = 1, \tag{3.126}$$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}\left[\sigma^{\otimes n} Q_n\right] = -H_r(\rho||\sigma), \tag{3.127}$$

where

$$Q_n := \sum_{(p,\rho') \in \mathcal{C}_r} T(np, \rho'). \tag{3.128}$$

Proof. See [16]. □

It should be pointed out that Theorem III.5.1 was proven nearly two decades after the first proposed quantum Sanov theorem Theorem III.4.1. When we compare the two results, the first obvious difference is that Theorem III.4.1 holds for open sets $\mathcal{A} \subset D(\mathcal{H})$, while Theorem III.5.1 only holds for when $\mathcal{A} = \{\rho\}$ is a singleton. The power of Theorem III.5.1 is that the sequence of tests Q_n are **independent** of the state σ, which is not the case in Theorem III.4.1. Another noteworthy observation is that the Hayashi relative entropy is bounded above the Umegaki relative entropy: $\forall \sigma, \rho \in D(\mathcal{H})$ and $r > 0$,

$$H_r(\rho||\sigma) \leq D_{KL}(\rho||\sigma). \tag{3.129}$$

In the next section we will discover that due to the independence of the sequence of tests Q_n, it is strictly necessary for (3.129) to hold.

Chapter IV

A universal quantum Sanov theorem for type $\mathbf{I}_d$

This chapter focuses on presenting the new results obtained during this internship. We are particularly interested in studying the role of the Umegaki relative entropy in the quantum Sanov theorem as proposed by [6, 23]. In the seminal paper [6], an example was constructed showing that in the qubit setting ($\mathcal{H} = \mathbb{C}^2$) it is generally not possible to construct a sequence of tests $(T_n)_{n\in\mathbb{N}}$ that separates a set $\mathcal{A} \subset D(\mathbb{C}^{\otimes 2})$ from an arbitrary collection of states $\mathcal{B}$ at rate equal to the Umegaki relative entropy. In the finite dimensional i.i.d setting, a sequence of tests (T_n) **separates** two disjoint sets of states $\mathcal{A}, \mathcal{B} \subseteq D(\mathcal{H})$ if for all $\rho \in \mathcal{A}$ and all $\sigma \in \mathcal{B}$ we have

$$\lim_{n\to\infty} \mathrm{Tr}[\rho^{\otimes n} T_n] = 1, \tag{4.1}$$

$$\lim_{n\to\infty} \mathrm{Tr}[\sigma^{\otimes n} T_n] = 0. \tag{4.2}$$

The qubit example in [6] showed that if a sequence of tests $(T_n)_{n\in\mathbb{N}}$ can separate a set $\mathcal{A} \subset D(\mathbb{C}^{\otimes 2})$ from a single state $\mathcal{B} = \{\sigma\} \subset D(\mathbb{C}^{\otimes 2})$ while simultaneously ensuring:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} T_n] = -D(\rho||\sigma), \tag{4.3}$$

then $T_n = T_n(\mathcal{A}, \mathcal{B})$ must depend on both the $\mathcal{A}$ and $\mathcal{B}$. The first new result we present in this section is a no-go type theorem (Theorem IV.1) which generalizes the example given in [6]. Namely, by extending an argument made in [28] we show that for sufficiently large Hilbert spaces ($\dim(\mathcal{H}) \geq 4$) there always exist states $\rho \in D(\mathcal{H})$ such that there is no sequence of tests $n \mapsto T_n \in D(\mathcal{H}^{\otimes n})$ that separates ρ from an arbitrary state $\sigma \in D(\mathcal{H})$ while ensuring (4.3) holds. This result showing the restriction in (4.3) is to strong of a condition, and function different from the Umegaki relative entropy should be considered. Our no-go theorem emphasizes the relevance of our second result stated in Theorem IV.2.2, which is a universal quantum Sanov theorem. While there have been several quantum extensions of the Sanov theorem, e.g., [6, 17, 23, 31], these results largely describe the optimal asymptotic behaviour of the type II error in quantum state discrimination when both the null and the alternative hypotheses are used to construct a sequence of separating tests. A notably distinct paper is [16], which addressed the problem of separating a state $\rho \in D(\mathcal{H})$ from an arbitrary $\sigma \in D(\mathcal{H})$ by constructing a quantum analogue of the empirical measure. However, this result does not go beyond

separating two single states.

The distinguishing features of our quantum Sanov theorem are that for a measurable set of states $\mathcal{A} \subseteq D(\mathcal{H})$ we construct a sequence of tests $T_n = T_n(\mathcal{A})$ only dependent on $\mathcal{A}$, such that $(T_n)_{n\in\mathbb{N}}$ universally separates $\mathcal{A}$ from any state $\sigma \in D(\mathcal{H})\backslash\mathcal{A}$, but necessarily with a rate function that is bounded above by the Umegaki relative entropy. Additionally, our quantum Sanov theorem is a direct analogue of the classical Sanov theorem, which also has the crucial universality feature we will soon define. The structure of this chapter is as follows. In Section IV.1 the concept of universality is discussed in the classical and quantum settings. The main goal of this section is to prove our no-go result stated in Theorem IV.1. To achieve this we extend two crucial results in [28]. In Section IV.2 we build the machinery needed to prove the central result of this paper, which is the universal quantum Sanov theorem stated in Lemma IV.2.1. Afterwards, in Corollary Theorem IV.2.2 we elaborate upon our findings and conclude a direct universal analogue of the quantum Sanov theorem proposed in [6].

IV.1 Universality

The universality feature of the classical Sanov Theorem II.2.2 is precisely the fact that $(L_n \in \Omega) = L_n^{-1}(\Omega)$ is a sequence of subsets that are independent of the prior q, and these subsets can separate any q from Ω with a rate function equal to the KL relative entropy. We define the quantum analogue of universality as follows.

Definition IV.1.1 (Universal Tests)**.** *Let $\mathcal{A} \subseteq D(\mathcal{H})$ and $I : D(\mathcal{H}) \times D(\mathcal{H}) \to [0,\infty]$. Then the sequence $n \mapsto T_n \in B(\mathcal{H}^{\otimes n})$ of operators satisfying $0 \leq T_n \leq \mathbf{1}$ with $T_n = T_n(\mathcal{A})$, is said to **universally** separate $\mathcal{A}$ at a rate $I : D(\mathcal{H}) \times D(\mathcal{H}) \to [0,\infty]$ if:*

$$\lim_{n\to\infty} \mathrm{Tr}[\rho^{\otimes n} T_n] = 1, \quad \forall \rho \in \mathcal{A}, \qquad \text{(Typicality)} \qquad (4.4)$$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} T_n] = -\inf_{\rho\in\mathcal{A}} I(\rho,\sigma), \quad \forall \sigma \in D(\mathcal{H}). \qquad \text{(Separation rate)} \qquad (4.5)$$

The first paper trying to extend the Sanov theorem into the (finite dimensional) quantum probability regime was [6]. Since then, the quantum Sanov theorem has been generalized by increasing the class of null and alternative hypotheses that one can consider, most recently in [23]. In Theorem III.4.1 we recall the first quantum Sanov theorem proved in [6], which we call a non-universal quantum Sanov Theorem III.4.1. For fixed a state $\sigma \in D(\mathcal{H})$ and a open collection of states $\mathcal{A} \subset D(\mathcal{H})$, there exists a sequence $\{P_n\}_{n\in\mathbb{N}}$ of projections $P_n \in B(\mathcal{H}^{\otimes n})$ such that

$$\lim_{n\to\infty} \mathrm{Tr}[\rho^{\otimes n} P_n] = 1, \quad \forall \rho \in \mathcal{A}, \qquad \text{(Typicality)} \qquad (4.6)$$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} P_n] = -\inf_{\rho\in\mathcal{A}} D(\rho||\sigma). \qquad \text{(Separation rate)} \qquad (4.7)$$

The non-universality of Theorem III.4.1 follows from the fact that the sequence of separating projections $(P_n)_{n\in\mathbb{N}}$ is dependent on the prior state σ. In Theorem IV.1 we prove that it is generally not possible to create a sequence of tests that universally separates

arbitrary states σ from a fixed state ρ at a rate equal to the Umegaki relative entropy. Namely, by extending a result from [28] we show that for a $d-$dimensional Hilbert space $\mathcal{H}$ where $d \geq 4$, there are always states $\rho \in D(\mathcal{H})$ that cannot be universally separated from arbitrary states σ at a rate function equal to the Umegaki relative entropy.

Theorem IV.1.1 (No-go Theorem)**.** *Let $d \geq 4$ and $\mathcal{H}$ a $d-$dimensional Hilbert space. Then for all $m \in \{1, \cdots, d\}$ there exists a $\rho \in D(\mathcal{H})$ with rank m such that there does not exist a sequence of tests $(T_n)_{n\in\mathbb{N}}$ that universally separates $\mathcal{A} = \{\rho\}$ at a rate equal to the Umegaki relative entropy. In other words, the sequence of projections in Theorem III.4.1 necessarily depends on σ.*

To prove we will extend [28, Lemma IV.2] to Lemma IV.1.1 below. First, we define two operator means. Let $\sigma_0, \sigma_1 \in B(\mathcal{H})$ be two invertible operators and $\lambda \in (0, 1)$. Then the $\lambda-$parametrized geometric mean and the $\lambda-$parametrized arithmetic mean are defined as follows,

$$\sigma_0 \#_\lambda \sigma_1 := \sigma_0^{-\frac{1}{2}} (\sigma_0^{\frac{1}{2}} \sigma_1 \sigma_0^{\frac{1}{2}})^\lambda \sigma_0^{-\frac{1}{2}}, \tag{4.8}$$

$$\sigma_0 \Delta_\lambda \sigma_1 := (1-\lambda)\sigma_0 + \lambda\sigma_1. \tag{4.9}$$

Lemma IV.1.1. *Let $\mathcal{K}$ be arbitrary finite dimensional Hilbert space with $\dim(\mathcal{K}) = d$. Assume $\sigma_0, \sigma_1 \in D(\mathcal{K})$ are two faithful states such that $[\sigma_0, \sigma_1] \neq 0$. Then there exists a state $\rho \in D(\mathcal{K})$ of arbitrary rank $1 \leq \operatorname{rank}(\rho) \leq d$ such that*

$$D(\rho||\sigma_0 \#_\lambda \sigma_1) < (1-\lambda) D(\rho||\sigma_0) + \lambda D(\rho||\sigma_1). \tag{4.10}$$

Proof. Here we use a similar argument as in [28, Lemma IV.2]. Define

$$\Gamma_\lambda(\sigma_0, \sigma_1) := \log(\sigma_0 \#_\lambda \sigma_1) - (1-\lambda)\log(\sigma_0) - \lambda \log(\sigma_1), \tag{4.11}$$

which is a self-adjoint operator. Then by [28, Lemma II.13] we know

$$\Gamma_\lambda(\sigma_0, \sigma_1) = 0 \iff [\sigma_0, \sigma_1] = 0. \tag{4.12}$$

Also for $A > 0$ we have $\operatorname{Tr}[\log A] = \log\det(A)$ and for $x \in \mathbb{R}\backslash\{0\}$ we have $\det(A^x) = \det(A)^x$. Making use of these simple properties we see that

$$\begin{aligned} \operatorname{Tr}[\Gamma_\lambda(\sigma_0, \sigma_1)] &= \log\det(\sigma_0 \#_\lambda \sigma_1) - (1-\lambda)\log\det(\sigma_0) - \lambda\log\det(\sigma_1) \\ &= \log[\det(\sigma_0)^{\lambda-1}\det(\sigma_1)^\lambda] - (1-\lambda)\log\det(\sigma_0) - \lambda\log\det(\sigma_1) \\ &= 0. \end{aligned} \tag{4.13}$$

Therefore since $\Gamma_\lambda(\sigma_0, \sigma_1)$ is self-adjoint we know the maximum eigenvalue μ_1 of Γ_λ satisfies $\mu_1 > 0$. Now consider the spectral decomposition of $\Gamma_\lambda(\sigma_0, \sigma_1)$

$$\Gamma_\lambda(\sigma_0, \sigma_1) = \sum_{i=1}^{d} \mu_i |\psi_i\rangle \langle\psi_i|, \tag{4.14}$$

where $\mu_1 \geq \cdots \geq \mu_d$. Since $\mu_1 > 0$ for any integer $m \in \{1, \cdots, d\}$, we know there exists a convex combination $\{\varepsilon_1, \cdots, \varepsilon_m\} \subset (0, 1)$ such that

$$\sum_{i=1}^{d} \mu_i \varepsilon_i > 0. \tag{4.15}$$

But then define

$$\rho_m := \sum_{i=1}^{m} \varepsilon_i \left|\psi_i\right\rangle \left\langle\psi_i\right|, \tag{4.16}$$

where $(|\psi_i\rangle)_{i=1}^m$ are the eigenvectors of $\Gamma_\lambda(\sigma_0, \sigma_1)$. Then

$$\mathrm{Tr}[\Gamma_\lambda(\sigma_0, \sigma_1)\rho_m] = \sum_{i=1}^{m} \mu_i \varepsilon_i > 0. \tag{4.17}$$

So for two invertible non commuting states σ_0, σ_1 and $\lambda \in (0,1)$ we have

$$\mathrm{Tr}[\Gamma_\lambda(\sigma_0, \sigma_1)\rho_m] > 0 \iff \mathrm{Tr}[\rho_m \log(\sigma_0 \#_\lambda \sigma_1)] > (1-\lambda)\mathrm{Tr}[\rho_m \log \sigma_0] + \lambda \mathrm{Tr}[\rho_m \log \sigma_1].$$

By subtracting

$$\mathrm{Tr}[\rho_m \log \rho_m] = (1-\lambda)\mathrm{Tr}[\rho_m \log \rho_m] + \lambda \mathrm{Tr}[\rho_m \log \rho_m], \tag{4.18}$$

from both sides we get the desired result. □

Proof of Theorem IV.1. To show this we alter the proof of [28, Theorem IV.3]. First assume that $\dim(\mathcal{H}) = d$ is even. Then, let $\mathcal{K}$ be a $\frac{d}{2}$-dimensional Hilbert space, then by assumption $\dim(\mathcal{K}) \geq 2$. Let $\sigma_0, \sigma_1 \in D(\mathcal{K})$ be two faithful states such that $[\sigma_0, \sigma_1] \neq 0$. Then for $\lambda \in (0,1)$ we have

$$0 \leq \sigma_0 \#_\lambda \sigma_1 \leq \sigma_0 \Delta_\lambda \sigma_1, \tag{4.19}$$

as shown in [21]. Furthermore, for any positive definite square matrices A, B, C, D,

$$(A \otimes C) \#_\lambda (B \otimes D) = (A \#_\lambda B) \otimes (C \#_\lambda D). \tag{4.20}$$

Then for all $n \in \mathbb{N}$ and tests $T_n \in B(\mathcal{K}^{\otimes n})$ we have

$$\mathrm{Tr}[(\sigma_0 \#_\lambda \sigma_1)^{\otimes n} T_n] = \mathrm{Tr}[(\sigma_0^{\otimes n} \#_\lambda \sigma_1^{\otimes n}) T_n] \leq (1-\lambda)\mathrm{Tr}[\sigma_0^{\otimes n} T_n] + \lambda \mathrm{Tr}[\sigma_1^{\otimes n} T_n]. \tag{4.21}$$

By Lemma IV.1.1 we can take a $\rho \in D(\mathcal{K})$ of arbitrary rank satisfying (4.10) for σ_0 and σ_1. Now we can proceed by a proof by contradiction. Assume that for all state $\rho \in D(\mathcal{H})$ there exists a sequence of tests $(T_n)_{n\in\mathbb{N}}$ such that for all $\sigma \in D(\mathcal{H})$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} T_n] = -D(\rho||\sigma). \tag{4.22}$$

Consider the states $\hat{\rho}, \hat{\sigma_0}, \hat{\sigma_1} \in D(\mathcal{H})$ defined as

$$\hat{\rho} = \frac{1}{2}\big(\rho \oplus \rho\big), \quad \hat{\sigma_0} = \frac{1}{2}\big(\sigma_0 \oplus \sigma_1\big), \quad \hat{\sigma_1} = \frac{1}{2}\big(\sigma_1 \oplus \sigma_0\big). \tag{4.23}$$

Then by invariance of the relative entropy under isometries we get

$$D(\hat{\rho}||\hat{\sigma_0}) = D(V^\dagger \hat{\rho} V||V^\dagger \hat{\sigma} V) = D(\hat{\rho}||\hat{\sigma_1}), \tag{4.24}$$

where we use the isometry $V(\sigma_0 \oplus \sigma_1) = \sigma_1 \oplus \sigma_0$. But by Lemma IV.1.1 for $\lambda \in (0,1)$ we know

$$D(\hat{\rho}||\hat{\sigma_0}) = D(\hat{\rho}||\hat{\sigma_1}) > D(\hat{\rho}||\hat{\sigma_0} \#_\lambda \hat{\sigma_1}). \tag{4.25}$$

Let $\beta = \mathrm{Tr}[\hat{\sigma}_0 \#_\lambda \hat{\sigma}_1] > 0$. By (4.21) we know that for all $n \in \mathbb{N}$ and tests $0 \leq T_n \leq \mathbf{1}$ the following holds:

$$\mathrm{Tr}[\beta^{-n}(\hat{\sigma}_0 \#_\lambda \hat{\sigma}_1)^{\otimes n} T_n] \leq \frac{1}{\beta^n} \max\{\mathrm{Tr}[\hat{\sigma}_0^{\otimes n} T_n], \mathrm{Tr}[\hat{\sigma}_1^{\otimes n} T_n]\}, \tag{4.26}$$

thus

$$\frac{1}{n} \log \mathrm{Tr}[\beta^{-n}(\hat{\sigma}_0 \#_\lambda \hat{\sigma}_1)^{\otimes n} T_n] \leq \log(\beta) + \max\left\{\frac{1}{n} \log \mathrm{Tr}[\hat{\sigma}_0^{\otimes n} T_n], \frac{1}{n} \log \mathrm{Tr}[\hat{\sigma}_1^{\otimes n} T_n]\right\}. \tag{4.27}$$

By assumption we can select $(T_n)_{n\in\mathbb{N}}$ such that in the limit of $n \to \infty$ we get

$$-D(\hat{\rho}||\beta^{-1}(\hat{\sigma}_0 \#_\lambda \hat{\sigma}_1)) \leq \log(\beta) + \max\{-D(\hat{\rho}||\hat{\sigma}_0), -D(\hat{\rho}||\hat{\sigma}_1)\}. \tag{4.28}$$

Upon using the relation

$$\begin{aligned} D(\hat{\rho}||\beta^{-1}(\hat{\sigma}_0 \#_\lambda \hat{\sigma}_1)) &= \mathrm{Tr}[\hat{\rho}(\log \hat{\rho} - \log \beta^{-1}(\hat{\sigma}_0 \#_\lambda \hat{\sigma}_1)) \\ &= D(\hat{\rho}||\hat{\sigma}_0 \#_\lambda \hat{\sigma}_1) + \log(\beta), \end{aligned} \tag{4.29}$$

which is generally true for ρ, σ_0, σ_1, we get

$$D(\hat{\rho}||\hat{\sigma}_0 \#_\lambda \hat{\sigma}_1) \geq D(\hat{\rho}||\hat{\sigma}_0) = \min\{D(\hat{\rho}||\hat{\sigma}_0), D(\hat{\rho}||\hat{\sigma}_1)\}. \tag{4.30}$$

However, this leads to a contradiction because by construction of ρ we had (4.25).

Now we treat the odd dimensional case. Assume that $\dim(\mathcal{H}) = 2d + 1$ for some integer $d \geq 2$. Then consider the $d-$dimensional Hilbert space $\mathcal{K}'$, by assumption $\dim(\mathcal{K}') \geq 2$. Then similar to before take two faithful states $\sigma_0, \sigma_1 \in D(\mathcal{K}')$, by Lemma IV.1.1 we can take a state $\rho \in D(\mathcal{K}')$ with arbitrary rank satisfying (4.10) for σ_0 and σ_1. Now consider the operators $\rho', \sigma_0', \sigma_1' \in D(\mathcal{H})$ defined as

$$\rho' := \frac{1}{3}\big(\rho \oplus 1 \oplus \rho\big), \quad \hat{\sigma_0} = \frac{1}{3}\big(\sigma_0 \oplus 1 \oplus \sigma_1\big), \quad \hat{\sigma_1} = \frac{1}{3}\big(\sigma_1 \oplus 1 \oplus \sigma_0\big). \tag{4.31}$$

Note that in a matrix representation this would look like

$$\sigma_0' = \frac{1}{3}\begin{pmatrix} \sigma_0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma_1 \end{pmatrix}, \tag{4.32}$$

where all zeros are suitable rectangular matrices of zeros. But, by invariance of the relative entropy under isometries we get

$$D(\rho'||\sigma_0') = D(\rho'||\sigma_1'). \tag{4.33}$$

With another proof by contradiction (as in the even case) we can identically show our assumption would imply

$$D(\rho'||\sigma_0' \#_\lambda \sigma_1') \geq \min\{D(\rho'||\sigma_0'), D(\rho'||\sigma_1')\}, \tag{4.34}$$

while we know $D(\rho'||\sigma_0') > D(\rho'||\sigma_0' \#_\lambda \sigma_1')$. And for $\rho \in D(\mathcal{H})$, there does not exist a sequence of tests $(T_n)_{n\in\mathbb{N}}$ such that for all $\sigma \in D(\mathcal{H})$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} T_n] = -D(\rho||\sigma). \tag{4.35}$$

□

IV.2 Universal quantum Sanov theorem

The crucial tool we will be making use of during this section is an informationally complete POVM.

Definition IV.2.1. *Let Ω be a finite set, and $\mu : \Omega \to B(\mathcal{H})$ be a POVM. We call μ an* ***information complete*** *POVM (IC-POVM), if for all $\sigma, \rho \in D(\mathcal{H})$ we have*

$$\rho = \sigma \iff \mathrm{Tr}[\mu(\omega)\rho] = \mathrm{Tr}[\mu(\omega)\sigma], \quad \forall \omega \in \Omega. \tag{4.36}$$

In the finite dimensional setting IC-POVM's always exist because the condition in (4.36) is equivalent to requiring $\mathrm{span}_{\mathbb{C}}\{\mu(\omega); \omega \in \Omega\} = B(\mathcal{H})$. The strength of an IC-POVM is highlighted in Proposition IV.2.1, where we show with an IC-POVM we are able to bijectively map all quantum states $D(\mathcal{H})$ to a collection $\Lambda \subseteq \mathrm{Prob}(A)$ of classical probability distributions.

Proposition IV.2.1. *Let $\mu : \Omega \to B(\mathcal{H})$ be an IC-POVM. Then there exists a closed convex set $\Lambda \subseteq Prob(\Omega)$ and a homeomorphism $\Phi : \Lambda \to D(\mathcal{H})$. We call the map Φ a* ***complete tomography*** *of $D(\mathcal{H})$.*

We will shortly provide a proof of this result. The significance of an IC-POVM is that we are now able to reduce any finite dimensional quantum system to a classical setting! Since quantum randomness is not the same as classical randomness, the *classical reduction* of a quantum system to a classical system comes at a cost. Consider an arbitrary complete tomography Φ defined by an IC-POVM μ as in Proposition IV.2.1. Then upon reducing all quantum states $D(\mathcal{H})$ to a collection of probability measure Λ, we have the following new entropy between two quantum states $\rho, \sigma \in D(\mathcal{H})$,

$$D_\Phi(\rho||\sigma) := D_{KL}(\Phi^{-1}(\rho)||\Phi^{-1}(\sigma)). \tag{4.37}$$

By the monotonicity of the Umegaki relative entropy Proposition III.1.3, if we consider the map Φ^{-1} is a quantum-to-classical reduction then

$$D_\Phi(\rho||\sigma) = D(\Phi^{-1}(\rho)||\Phi^{-1}(\sigma)) \leq D(\rho||\sigma), \tag{4.38}$$

where the inequality is strict if $[\rho, \sigma] \neq 0$. Hence, for two non-commuting quantum states $\rho, \sigma \in D(\mathcal{H})$,

$$0 < D_\Phi(\rho||\sigma) < D(\rho||\sigma), \tag{4.39}$$

and the complete tomography reduces the distinguishability of ρ from σ, however it does preserves the fact that $\rho \neq \sigma$. The preservation of the distinguishability of two quantum states is exactly what makes an IC-POVM useful for developing our universal quantum Sanov theorem.

Proof of Proposition IV.2.1. We begin by defining

$$\Lambda := \{p \in \mathrm{Prob}(\Omega); \exists \rho \in D(\mathcal{H}) \text{ s.t } \forall \omega \in \Omega, \quad p(\omega) = \mathrm{Tr}[\rho\mu(\omega)]\}. \tag{4.40}$$

Because μ is an IC-POVM, for each $p \in \Lambda$ there exists only one $\rho \in D(\mathcal{H})$ such that

$$p(\omega) = \mathrm{Tr}[\rho\mu(\omega)], \quad \forall \omega \in \Omega. \tag{4.41}$$

Define the function $\Phi : \Lambda \to D(\mathcal{H})$ which maps each $p \in \Lambda$ to its unique state $\rho \in D(\mathcal{H})$ satisfying (4.41). Clearly Φ is injective and surjectivity easily follows because for any $\sigma \in D(\mathcal{H})$ we can define $q \in \Lambda$ as

$$q(\omega) := \mathrm{Tr}[\mu(\omega)\sigma], \quad \forall \omega \in \Omega. \tag{4.42}$$

Because Φ is bijective it has a well defined inverse $\Phi^{-1} : D(\mathcal{H}) \to \Lambda$, which we will now show is continuous. Let $(\rho_n)_{n\in\mathbb{N}} \subseteq D(\mathcal{H})$ be a sequence of density operators converging to some $\rho_0 \in D(\mathcal{H})$. Then for all $n \in \mathbb{N}$ each ρ_n has a $p_n \in \Lambda$ given by

$$p_n = \Phi^{-1}(\rho_n), \tag{4.43}$$

or, stated otherwise, each $p_n \in \Lambda$ satisfies

$$p_n(\omega) = \mathrm{Tr}[\rho_n \mu(\omega)], \quad \forall \omega \in \Omega. \tag{4.44}$$

Then by continuity of the trace in finite dimensions we get for all $\omega \in \Omega$,

$$\lim_{n\to\infty} \Phi^{-1}(\rho_n)(\omega) = \lim_{n\to\infty} \mathrm{Tr}[\rho_n \mu(\omega)] = \mathrm{Tr}[\rho_0 \mu(\omega)] = \Phi^{-1}(\rho_0)(\omega), \tag{4.45}$$

and because Λ is a subset of some finite dimensional vector space, point-wise convergence of $\Phi^{-1}(\rho_n)$ to $\Phi^{-1}(\rho_0)$ implies convergence with respect to any norm on Λ thus,

$$\lim_{n\to\infty} \Phi^{-1}(\rho_n) = \Phi^{-1}(\rho_0), \tag{4.46}$$

and Φ^{-1} is bijective and continuous. Moving on, the set of positive operators $B(\mathcal{H})^+$ is a closed cone, and by continuity of the trace in finite dimensions $D(\mathcal{H})$ is closed and bounded, and therefore compact by the Heine-Borel theorem. Therefore, Φ^{-1} is a continuous bijection from a compact metric space $D(\mathcal{H})$ to another metric space Λ and by [34, Theorem 4.17] $(\Phi^{-1})^{-1} = \Phi$ is continuous. So we have shown that Φ is a homeomorphism and we need only prove Λ is a closed convex set. Note that $\Lambda = \Phi^{-1}(D(\mathcal{H}))$ is closed because $D(\mathcal{H})$ is closed and Φ^{-1} is continuous. Finally, let $p, q \in \Lambda$ and set $\rho = \Phi(p), \sigma = \Phi(q)$. Then as $D(\mathcal{H})$ is a convex set we have for all $\lambda \in [0, 1]$,

$$\rho_\lambda = \lambda\rho + (1 - \lambda)\sigma, \tag{4.47}$$

and for all $\omega \in \Omega$,

$$\Phi^{-1}(\rho_\lambda)(\omega) = \lambda \mathrm{Tr}[\rho\mu(\omega)] + (1 - \lambda)\mathrm{Tr}[\sigma\mu(\omega)] = \lambda p(\omega) + (1 - \lambda)q(\omega). \tag{4.48}$$

So clearly $\Phi^{-1}(\rho_\lambda) = \lambda p + (1 - \lambda)q \in \Lambda$. □

With an IC-POVM we are able to construct a quantum variant to the empirical measure, much like the goal of [16]. Given an IC-POVM $\mu : \Omega \to B(\mathcal{H})$ we can construct a sequence of POVM's $\mu_n : \Omega^n \to B(\mathcal{H}^{\otimes n})$ as follows. Let $n \in \mathbb{N}$, and for $(\omega_1, \cdots, \omega_n) \in \Omega^n$ define

$$\mu_n(\omega_1, \cdots, \omega_n) := \mu(\omega_1) \otimes \cdots \otimes \mu(\omega_n). \tag{4.49}$$

In general for positive matrices A, B, C, D if $0 \leq A \leq B$ and $0 \leq C \leq D$ then $0 \leq A \otimes C \leq B \otimes D$. So for all $(\omega_1, \cdots, \omega_n) \in \Omega^n$ we have

$$0 \leq \mu(\omega_1) \otimes \cdots \otimes \mu(\omega_n) \leq \mathbf{1}^{\otimes n}, \tag{4.50}$$

where $\mathbf{1}^{\otimes n}$ is the n-fold tensor product of the identity $\mathbf{1} \in B(\mathcal{H})$. Additionally, by linearity of the tensor product

$$\sum_{\omega_1 \in \Omega} \cdots \sum_{\omega_n \in \Omega} \mu(\omega_1) \otimes \cdots \otimes \mu(\omega_n) = \Big(\sum_{\omega_1 \in \Omega} \mu(\omega_1) \Big) \otimes \cdots \otimes \Big(\sum_{\omega_n \in \Omega} \mu(\omega_n) \Big) = \mathbf{1}^{\otimes n}, \quad (4.51)$$

and for any subset $G \subseteq \Omega^n$ we define

$$\mu_n(G) := \sum_{(\omega_1, \cdots, \omega_n) \in G} \mu_n(\omega_1, \cdots, \omega_n). \quad (4.52)$$

Let $\rho \in D(\mathcal{H})$, and consider the function Φ defined in Proposition IV.2.1. Fix $p = \Phi^{-1}(\rho)$, then for $n \in \mathbb{N}$, let p^n denote the Bernoulli measure on Ω^n induced by p. For any $G \subseteq \Omega^n$ we have,

$$\begin{aligned} p^n(G) &= \sum_{(\omega_1, \cdots, \omega_n) \in G} p(\omega_1) \cdots p(\omega_n) = \sum_{(\omega_1, \cdots, \omega_n) \in G} \mathrm{Tr}[\rho \mu(\omega_1)] \cdots \mathrm{Tr}[\rho \mu(\omega_n)] \\ &= \sum_{(\omega_1, \cdots, \omega_n) \in G} \mathrm{Tr}[\rho^{\otimes n} \mu_n(\omega_1, \cdots, \omega_n)] = \mathrm{Tr}[\rho^{\otimes n} \mu_n(G)]. \end{aligned} \quad (4.53)$$

Lemma IV.2.1. *Let $\mu : \Omega \to B(\mathcal{H})$ be an IC-POVM and $\mathcal{A} \subseteq D(\mathcal{H})$ be a measurable set. For all $n \in \mathbb{N}$, define the sequence of tests $T_n := \mu_n(L_n^{-1}(\Phi^{-1}(\mathcal{A})))$, where $L_n : \Omega^n \to Prob(\Omega)$ is the empirical measure, μ_n is defined as in (4.52), and $\Phi : \Lambda \to D(\mathcal{H})$ is as in Proposition IV.2.1. Let $D_\Phi : D(\mathcal{H}) \times D(\mathcal{H}) \to [0, \infty]$ be defined as in (4.37). Then for all $\sigma \in D(\mathcal{H})$,*

$$\begin{aligned} -\inf_{\rho \in \mathcal{A}^\circ} D_\Phi(\rho || \sigma) &\leq \liminf_{n \to \infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} T_n] \\ &\leq \limsup_{n \to \infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} T_n] \\ &\leq -\inf_{\rho \in \mathcal{A}^-} D_\Phi(\rho || \sigma), \end{aligned} \quad (4.54)$$

where $\mathcal{A}^\circ, \mathcal{A}^-$ are the interior and closure of $\mathcal{A}$ respectively.

Proof. Let $G = \Phi^{-1}(\mathcal{A}) \subseteq \Lambda \subseteq \mathrm{Prob}(\Omega)$, then as Φ is continuous and $\mathcal{A}$ measurable, we know G is measurable. Now let $q = \Phi^{-1}(\sigma)$, then by Theorem II.2.2,

$$\begin{aligned} -\inf_{p \in G^\circ} D_{KL}(p || q) &\leq \liminf_{n \to \infty} \frac{1}{n} \log q^n(L_n \in G) \\ &\leq \limsup_{n \to \infty} \frac{1}{n} \log q^n(L_n \in G) \leq -\inf_{p \in G^-} D_{KL}(p || q). \end{aligned} \quad (4.55)$$

By (4.53) we can substitute

$$q^n(L_n \in G) = q^n(L_n^{-1} \circ \Phi^{-1}(\mathcal{A})) = \mathrm{Tr}[\sigma^{\otimes n} T_n], \quad (4.56)$$

and because Φ is a bijection for any $\mathcal{Q} \subseteq D(\mathcal{H})$

$$\inf_{\rho \in \mathcal{Q}} D(\Phi^{-1}(\rho) || \Phi^{-1}(\sigma)) = \inf_{p \in \Phi^{-1}(\mathcal{Q})} D(p || q). \quad (4.57)$$

□

We will now elaborate on how Lemma IV.2.1 can be viewed as a quantum extension of Sanov's theorem. The first remark is that Lemma IV.2.1 holds for arbitrary *measurable* sets, as opposed to only open sets $\mathcal{A} \subset D(\mathcal{H})$ in Theorem III.4.1, and singletons $\mathcal{A} = \{\rho\}$ in Theorem III.5.1. More interestingly, this result provides a very natural quantum version of the empirical measure. Namely, for an IC-POVM $\mu : \Omega \to B(\mathcal{H})$, we consider the measurement process of the POVM $\mu_n : \Omega^n \to D(\mathcal{H})$; which is the n-fold tensor product of μ. Just like in the classical setting, we can take n-measurements on i.i.d copies of σ (with μ), which produces a sequence $\omega = (\omega_1, \cdots, \omega_n) \in \Omega^n$. Using the classical empirical measure $L_n : \Omega^n \to \Lambda \subset \text{Prob}(\Omega)$, we can map the string of observations to a classical distribution $L_n(\omega) \in \text{Prob}(\Lambda)$. The key ingredient for turning the classical empirical measure into a quantum empirical measure is the homeomorphic map $\Phi : \Lambda \to D(\mathcal{H})$.

Definition IV.2.2 (Quantum Empirical Measure)**.** *Let $\mu : \Omega \to D(\mathcal{H})$ be an IC-POVM. Let $\Phi : \Lambda \to D(\mathcal{H})$ be the induced complete tomography by μ (Proposition IV.2.1). Then the **quantum empirical measure** $\hat{L_n} : \Omega^n \to D(\mathcal{H})$, is defined as*

$$\hat{L_n^{\mu}}(\omega_1, \cdots, \omega_n) := \Phi(L_n(\omega_1, \cdots, \omega_n)). \tag{4.58}$$

Just as in the classical case we can also consider $\hat{L_n^{\mu}} : \Omega^{\mathbb{N}} \to D(\mathcal{H})$, where for all $n \in \mathbb{N}$, $\hat{L_n^{\mu}}$ is a function of the first n-components in a sequence $\omega \in \Omega^{\mathbb{N}}$.

Since Φ is a homeomorphism, if $\omega \in \Omega^{\mathbb{N}}$ and $\lim_{n\to\infty} L_n(\omega) = q$ for $q \in \Lambda \subseteq \text{Prob}(\Omega)$, then

$$\lim_{n\to\infty} \hat{L_n^{\mu}}(\omega) = \Phi(q). \tag{4.59}$$

We can now state a reformulation of our universal quantum Sanov theorem matching the classical result more closely. Let $\mathcal{A} \subset D(\mathcal{H})$, moving forward we use the following notation:

$$(\hat{L_n^{\mu}} \in \mathcal{A}) := \hat{L_n^{\mu}}^{-1}(\mathcal{A}) \subset \Omega^n. \tag{4.60}$$

Let $\sigma \in D(\mathcal{H})$ and $q = \Phi^{-1}(\sigma)$, in the context of measuring n-copies of σ with the POVM $\mu : \Omega \to D(\mathcal{H})$, the probability that $\hat{L}_n^{\omega}(\omega) \in \mathcal{A}$, is given by

$$q^n(\hat{L_n^{\mu}} \in \mathcal{A}) = \text{Tr}[\sigma^{\otimes n} \mu_n(\hat{L_n^{\mu}} \in \mathcal{A})]. \tag{4.61}$$

Theorem IV.2.1 (A universal quantum Sanov theorem)**.** *Let $\mu : \Omega \to B(\mathcal{H})$ be an IC-POVM. Let Φ be the induced complete tomography by μ (Proposition IV.2.1), and μ_n be the n-fold tensor product of μ. Let $\hat{L}_n^{\mu} : \Omega^n \to D(\mathcal{H})$ be the quantum empirical measure as defined in Definition IV.2.2, and D_Φ be defined as in* (4.37)*. Then for all $\sigma \in D(\mathcal{H})$ and measurable sets $\mathcal{A} \subset D(\mathcal{H})$,*

$$\begin{aligned} -\inf_{\rho\in\mathcal{A}^\circ} D_\Phi(\rho||\sigma) &\leq \liminf_{n\to\infty} \frac{1}{n} \log \text{Tr}[\sigma^{\otimes n} \mu_n(\hat{L}_n^{\mu} \in \mathcal{A})] \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log \text{Tr}[\sigma^{\otimes n} \mu_n(\hat{L}_n^{\mu} \in \mathcal{A})] \\ &\leq -\inf_{\rho\in\mathcal{A}^-} D_\Phi(\rho||\sigma), \end{aligned} \tag{4.62}$$

where $\mathcal{A}^\circ, \mathcal{A}^-$ are the interior and closure of $\mathcal{A}$ respectively.

Proof. This result follows from Lemma IV.2.1. □

We will now restate our result in the traditional language of quantum hypothesis testing. Combining the results of Theorem IV.1, and Lemma IV.2.1 we see that there do exist sequences $(T_n)_{n\in\mathbb{N}}$ of tests that universally separate $\mathcal{A}$ at some rate I, but necessarily the rate function I cannot be equal to the Umegaki relative entropy. In the case that an IC-POVM is used the rate function is bounded above by the Umegaki relative entropy. We can summarize our observations as follows.

Theorem IV.2.2. *Let $\mu : \Omega \to B(\mathcal{H})$ be an IC-POVM and $\mathcal{A} \subseteq D(\mathcal{H})$ be open. Suppose $T_n = T_n(\mathcal{A})$ is defined as in Lemma IV.2.1 and $D_\Phi : D(\mathcal{H}) \times D(\mathcal{H}) \to [0,\infty]$ is defined as in* (4.37)*. Then*

$$\lim_{n\to\infty} \mathrm{Tr}[\rho^{\otimes n} T_n] = 1, \quad \forall \rho \in \mathcal{A}, \qquad \text{(Typicality)} \tag{4.63}$$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{Tr}[\sigma^{\otimes n} T_n] = -\inf_{\rho\in\mathcal{A}} D_\Phi(\rho||\sigma), \quad \forall \sigma \in D(\mathcal{H}). \qquad \text{(Separation rate)} \tag{4.64}$$

Proof. Since $\mathcal{A}$ is open we have,

$$-\inf_{\rho\in\mathcal{A}^\circ} D_{KL}(\Phi^{-1}(\rho)||\Phi^{-1}(\sigma)) = -\inf_{\rho\in\mathcal{A}^-} D_{KL}(\Phi^{-1}(\rho)||\Phi^{-1}(\sigma)). \tag{4.65}$$

Let $p = \Phi^{-1}(\rho)$ and $G = \Phi^{-1}(\mathcal{A})$. Then by [26, Proposition 4.1] we know

$$\lim_{n\to\infty} p^n(L_n \in G) = 1. \tag{4.66}$$

The rest of the statement follows by using (4.53). □

IV.2.1 An example

Here we provide an example of an IC-POVM $\mu : \Omega \to D(\mathcal{H})$, and the relative entropy $D_\Phi : D(\mathcal{H}) \times D(\mathcal{H}) \to [0,\infty]$.

Proposition IV.2.2. *Let $\mathcal{H}$ be a d-dimensional Hilbert space. If $I = \{1, \cdots, d+1\}$, then there exists a collection of distinct orthonormal bases $(\mathcal{B}_i)_{i\in I}$ such that if $\Omega := \cup_{i\in I}\mathcal{B}_i$, then the POVM $\mu : \Omega \to B(\mathcal{H})$ defined as*

$$\mu(|x\rangle) := \frac{1}{|\Omega|} |x\rangle \langle x|, \tag{4.67}$$

is an IC-POVM.

See Proposition A.4.1 in the Appendix for a sketch of the proof. For each basis $\mathcal{B}_i$ we construct the pinching map $V_i : D(\mathcal{H}) \to D(\mathcal{H})$ as:

$$V_i(\sigma) := \sum_{|x\rangle\in\mathcal{B}_i} \lambda_x |x\rangle \langle x|, \qquad \lambda_x := \langle x| \sigma |x\rangle. \tag{4.68}$$

Then the relative entropy $D_\Phi : D(\mathcal{H}) \times D(\mathcal{H})$ associated to μ is given by

$$D_\Phi(\rho||\sigma) = \frac{1}{|\Omega|} \sum_{i\in I} D(V_i(\rho)||V_i(\sigma)), \tag{4.69}$$

which we confirm with the following calculation. Let $\rho, \sigma \in D(\mathcal{H})$ and $p, q \in \text{Prob}(\Omega)$ such that $\Phi^{-1}(\rho) = p$ and $\Phi^{-1}(\sigma) = q$, where Φ is defined in Proposition IV.2.1. If $m = |\Omega|$ then,

$$D_{KL}(p||q) = \sum_{a\in\Omega} p(a)(\log p(a) - \log q(a)) \tag{4.70}$$

$$= \sum_{a\in\Omega} \frac{1}{m} \langle a| \rho |a\rangle \left(\log \frac{1}{m} \langle a| \rho |a\rangle - \log \frac{1}{m} \langle a| \sigma |a\rangle \right) \tag{4.71}$$

$$= \frac{1}{m} \sum_{i=1}^{m} \sum_{|x\rangle\in\mathcal{B}_i} \langle x| \rho |x\rangle (\log \langle x| \rho |x\rangle - \log \langle x| \sigma |x\rangle) \tag{4.72}$$

$$= \frac{1}{m} \sum_{i=1}^{m} \text{Tr}[V_i(\rho)(\log V_i(\rho) - \log V_i(\sigma))] = \frac{1}{m} \sum_{i=1}^{m} D(V_i(\rho)||V_i(\sigma)). \tag{4.73}$$

Chapter V

A universal quantum Sanov theorem in more generality

This chapter is dedicated to extending the universal quantum Sanov Lemma IV.2.1 beyond the type I_d setting for $d \in \mathbb{N}$. In the finite dimensional setting, the main feature of an IC-POVM $\mu : \Omega \to B(\mathcal{H})^+$, is that one can use it to separate any two states $\rho, \sigma \in D(\mathcal{H})$, namely

$$\rho = \sigma \iff \mathrm{Tr}[\rho\mu(\omega)] = \mathrm{Tr}[\sigma\mu(\omega)], \quad \forall \omega \in \Omega. \tag{5.1}$$

When $\mathcal{H}$ is a finite dimensional Hilbert space, we were able to reduce all quantum states $D(\mathcal{H})$ to classical probability distributions $\Lambda \subseteq \mathrm{Prob}(\Omega)$ with a IC-POVM. We will follow a very similar process to extend the universal quantum Sanov theorem. However, for infinite dimensional von Neumann algebras, not all IC-POVM's are suitable for reducing all normal states $\mathcal{S}_n(\mathcal{M})$ to a set of classical probability distributions. One of the advantages of working in the finite dimensional setting is that all operator topologies coincide and, there are sets of separating operators $\{\mu(\omega); \omega \in \Omega\}$ which have finitely many elements. Obviously, for more general von Neumann algebras the operator topologies never coincide and there is no finite collection of operators $\{\mu(\omega); \omega \in \Omega\}$ that separate all normal states. These two facts will require us to use the more general Sanov theorem presented for Polish spaces and arbitrary measure spaces in Section II.3.

The main result of this chapter will be proved in Section V.3, where we extend the universal quantum Sanov theorem to the type I_∞ case. A key part of our proof relies on the unique properties of the compact operators $K(\mathcal{H}) \subseteq B(\mathcal{H})$ (Definition B.0.2). In Section V.1, we discuss some preliminary material, such as the generalization of an IC-POVM to an arbitrary von Neumann algebra and a brief overview of the classification of factors. In Section V.2.1, we present a very generic construction of an IC-POVM that can be used to separate two distinct quantum states $\psi, \phi \in \mathcal{S}_n(\mathcal{M})$, much like in (5.1). Despite the promising construction, we present a example illustrating that the proof done in Section V.3 does not extend nicely beyond the type I_∞ case.[1] Details of their contributions are provided in the acknowledgement in Section I.2.

[1] This part of the project was contributed to by Stefaan Vaes and Martijn Caspers.

V.1 Preliminaries

We begin by by stating several useful definitions and results regarding von Neumann algebras.

Theorem V.1.1 ($(\mathcal{M}_*)^* \simeq \mathcal{M}$)**.** *Let $\mathcal{M} \subseteq B(\mathcal{H})$ be a von Neumann algebra and $\mathcal{M}_*$ its predual. Then $\mathcal{M}_* \subseteq \mathcal{M}^*$ is a Banach space when equipped with the norm on $\mathcal{M}_*$. Moreover, the Banach dual of $\mathcal{M}_*$ is isometrically isomorphic to $\mathcal{M}$, i.e. $(\mathcal{M}_*)^* \simeq \mathcal{M}$.*

Proof. See [7, Theorem III.2.4.1]. □

Now we will define a generalized notion of a POVM for arbitrary von Neumann algebras. We will closely follow Chapter 18 in [9].

Definition V.1.1 (General POVM)**.** *Take any Hilbert space $\mathcal{H}$ and consider a von Neumann algebra $\mathcal{M} \subseteq B(\mathcal{H})$. Let (X, Ω) be some measure space. Then a function $\mu : \Omega \to \mathcal{M}^+$, where $\mathcal{M}^+ = \{a \geq 0, a \in \mathcal{M}\}$, is a **positive operator valued measure** if the following conditions hold:*

1. *For all $A \in \Omega$, we have $\mu(A) \in \mathcal{M}^+$ while $\mu(X) = \mathbf{1}$ and $\mu(\emptyset) = 0$.*
2. *For a disjoint countable collection $(A_n)_{n\in\mathbb{N}} \subseteq \Omega$ we have*

$$\sum_{n}^{\infty} \mu(A_n) = \mu\Big(\bigcup_{n=1}^{\infty} A_n\Big), \tag{5.2}$$

where the limit of the sum in the left hand side is with respect to the weak operator or strong operator topology (Definition B.0.8).

Definition V.1.2 (Tensor product of von Neumann algebras)**.** *Let $\mathcal{M}$ and $\mathcal{N}$ be two von Neumann algebras. Consider*

$$\mathcal{M} \odot \mathcal{N} = \mathrm{span}_{\mathbb{C}}\{a \otimes b; a \in \mathcal{M}, b \in \mathcal{N}\}. \tag{5.3}$$

With $\mathcal{M} \otimes \mathcal{N}$ we denote the weak operator topology (Definition B.0.8) closure of $\mathcal{M} \odot \mathcal{N}$.

Proposition V.1.1. *The tensor product of two POVM's $\mu_1 : \Omega_1 \to \mathcal{M}$ and $\mu_2 : \Omega_2 \to \mathcal{N}$ extends to a POVM $\mu_1 \otimes \mu_2 : \Omega_{12} \to \mathcal{M} \otimes \mathcal{N}$, where Ω_{12} is the product σ-algebra of Ω_1 and Ω_2.*

Proof. [35, Proposition IV.4.23]. □

Definition V.1.3 (IC-POVM)**.** *In the same setting as Definition V.1.1, we say a POVM $\mu : \Omega \to \mathcal{M}^+$ is **information complete**, if for any two normal functionals φ and $\tilde{\varphi} \in \mathcal{M}_*$, we have*

$$\varphi = \tilde{\varphi} \iff \forall A \in \Omega, \varphi(\mu(A)) = \tilde{\varphi}(\mu(A)). \tag{5.4}$$

Proposition V.1.2. *Let (X, Ω) be a measure space and suppose*

$$\mu : \Omega \to \mathcal{M}^+,$$

is a positive operator valued measure. Then

$$\mu \text{ is information complete} \iff \{\mathrm{span}\{\mu(A); A \in \Omega\}\}^{-\sigma w} = \mathcal{M}. \tag{5.5}$$

Proof. [9, Proposition 18.1]. □

Definition V.1.4 (Factor)**.** *Let $\mathcal{M}$ be a von Neumann algebra. Define the* ***center*** *of $\mathcal{M}$ as*

$$Z(\mathcal{M}) = \mathcal{M} \cap \mathcal{M}'.$$

We call $\mathcal{M}$ a ***factor*** *if $Z(\mathcal{M}) = \mathbb{C}\mathbf{1}$.*

We can now state the classification of factors. It is important to remember that von Neumann algebras always have projections. In fact, by Theorem B.0.8 we have that the norm-closed linear span of all projections in a von Neumann algebra $\mathcal{M}$, is equal to $\mathcal{M}$.

Definition V.1.5. *Take two projections $p, q \in \mathcal{M}$. We say p is equivalent to q, and write $p \sim q$, if there exists a $u \in \mathcal{M}$ such that $p = uu^*$ and $q = u^*u$.*

Definition V.1.6. *A projection $p \in \mathcal{M}$ is called* ***finite*** *if for all projections $q \in \mathcal{M}$ satisfying $q \leq p$ and $q \sim p$, implies $p = q$. The projection p is called* ***minimal*** *if for all projections $q \in \mathcal{M}$ satisfying $q \leq p$ implies $q = 0$ or $q = p$. Moreover, a von Neumann algebra $\mathcal{M}$ is called* ***finite*** *if the identity $\mathbf{1} \in \mathcal{M}$ is a finite projection,* ***semi-finite*** *if for all finite projections $p \in \mathcal{M}$ we have $p \leq \mathbf{1}$, and* ***purely infinite*** *if all non-zero projections are not finite.*

Definition V.1.7. *Let $\mathcal{M} \subseteq B(\mathcal{H})$ be a factor.*

- *A factor $\mathcal{M}$ is* ***Type I*** *if it has at least one minimal projection $p \in \mathcal{M}$. Furthermore a type I factor is either:*
  - *Type I_d if $\mathbf{1}$ can be expressed as a linear combination of $d \in \mathbb{N}$ minimal projections in $\mathcal{M}$,*
  - *Type I_∞ if $\mathcal{M}$ is semi-finite and not finite.*
- *A factor $\mathcal{M}$ is* ***Type II*** *if it has no minimal projections, and at least one non-zero finite projection. Furthermore a type II factor is either:*
  - *Type I_1 if $\mathcal{M}$ is finite,*
  - *Type I_∞ if $\mathcal{M}$ is semi-finite and not finite.*
- *A factor $\mathcal{M}$ is* ***Type III*** *if all non-zero projections are infinite.*

Proposition V.1.3. *A factor $\mathcal{M}$ is either type I_d, type I_∞, type II_1, type II_∞, or type III.*

Proof. See [35, Corollary V.1.20]. □

V.2 Information complete operator valued measures

V.2.1 Constructed from predual

In this section we construct an information complete POVM for an arbitrary von Neumann algebra $\mathcal{M} \subseteq B(\mathcal{H})$, where $\mathcal{H}$ is separable. Or equivalently, if we equip the predual $\mathcal{M}_* \subset \mathcal{M}^*$, with the norm on $\mathcal{M}^*$, we assume $\mathcal{M}_*$ is norm-separable. It is worthwhile

to mention there is a very rich collection of von Neumann algebras with a separable predual, in fact one can construct von Neumann algebras of type II and III, see [7, III.3.4.3,IV.2.7.4] for more details. The first step we take in constructing an IC-POVM is to find a collection of operators $(\tilde{a}_{i,j})_{i,j\in\mathbb{N}} \subseteq \mathcal{M}^+$, that can separate any two non-equal normal states $\varphi_i, \varphi_j \in \mathcal{M}_*$.

Lemma V.2.1. *Let $\mathcal{M}$ be a von Neumann algebra and $(\varphi_n)_{n\in\mathbb{N}} \subseteq \mathcal{M}_* \subset \mathcal{M}^*$ be a norm dense subset of $\mathcal{M}_*$, where $\mathcal{M}_*$ inherits the norm from $\mathcal{M}^*$. Assume without loss of generality for all $i, j \in \mathbb{N}$, $\varphi_i \neq \varphi_j$ if $i \neq j$. Let $i, j \in \mathbb{N}$ such that $i \neq j$, then there exists a positive operator $\tilde{a}_{i,j} \in \mathcal{M}^+$, with $||\tilde{a}_{i,j}|| = 1$, satisfying*

$$0 < \frac{1}{5}||\varphi_i - \varphi_j|| \leq |\varphi_i(\tilde{a}_{i,j}) - \varphi_j(\tilde{a}_{i,j})|. \tag{5.6}$$

Proof. For any $\tau \in \mathcal{M}_*$,

$$\begin{aligned}
||\tau|| =& \sup_{||a||=1} |\tau(a)| \\
=& \sup\{|\tau(u_1) - \tau(v_1) + i(\tau(u_2) - \tau(v_2))|; ||u_1 - v_1 + i(u_2 - v_2)|| = 1, u_1, u_2, v_1, v_2 \in \mathcal{M}^+\} \\
\leq& 4 \cdot \sup_{a\geq 0, ||a||=1} |\tau(a)|,
\end{aligned} \tag{5.7}$$

where we make use of the polar identity for an operator $a \in \mathcal{M}$, namely, we can always decompose $a = u_1 - v_1 + i(u_2 - v_2)$, into a linear combination of four positive operators $u_1, v_1, u_2, v_2 \in \mathcal{M}^+$. Now let $\tau = \varphi_i - \varphi_j$, and since $\varphi_i \neq \varphi_j$ we know $||\tau|| > 0$. By definition of the supremum, there exists a positive operator $\tilde{a}_{i,j} \in \mathcal{M}^+$ with norm $||\tilde{a}_{i,j}|| = 1$, such that

$$|\tau(\tilde{a}_{i,j})| \geq \frac{4}{5} \cdot \sup_{a\geq 0, ||a||=1} |\tau(a)|. \tag{5.8}$$

Fix such an $\tilde{a}_{i,j}$. Then

$$||\tau|| \leq 4 \cdot \sup_{a\geq 0, ||a||=1} |\tau(a)| \leq 4 \cdot \frac{5}{4}|\tau(\tilde{a}_{i,j})| \implies \frac{1}{5}||\tau|| \leq |\tau(\tilde{a}_{i,j})|. \tag{5.9}$$

□

Let $i, j \in \mathbb{N}$, such that $i \neq j$. Take an operator $\tilde{a}_{i,j} \in \mathcal{M}^+$ satisfying (5.6) and $||\tilde{a}_{i,j}|| = 1$. We will construct a new collection $(a_{i,j})_{i,j\in\mathbb{N}} \subseteq \mathcal{M}^+$ of operators using $(\tilde{a}_{i,j})_{i,j\in\mathbb{N}}$. The set $(a_{i,j})_{i,j\in\mathbb{N}} \subseteq \mathcal{M}^+$ will be used to construct an IC-POVM. Fix a $p \in \text{Prob}(\mathbb{N} \times \mathbb{N})$ such that for all $i \neq j$ we have $p(i, j) > 0$, and $p(i, i) = 0$ if $i \geq 1$. Then define,

$$\begin{aligned}
a_{i,j} =& p(i, j)\tilde{a}_{i,j} \qquad \text{for } i \neq j, \\
a_{i,i} =& 0 \qquad\qquad\qquad\quad \text{for } i \geq 1, \\
a_{0,0} =& \mathbf{1} - \sum_{i,j=0,i\neq j}^{\infty} a_{i,j}.
\end{aligned} \tag{5.10}$$

Note, the infinite sum in the definition of $a_{0,0}$ converges, because $\mathcal{M}$ is a Banach space, and the sum converges absolutely in the norm topology,

$$\sum_{i,j=0,i\neq j}^{\infty} ||a_{i,j}|| = \sum_{i,j=0,i\neq j} p(i, j)||\tilde{a_{i,j}}|| = 1 - p(0, 0). \tag{5.11}$$

Moreover, $\sum_{i,j=0,i\neq j}^{\infty} a_{i,j}$ is a positive operator since $\mathcal{M}^+$ is a norm-closed cone. We will now show that the countable subset $(a_{i,j})_{i,j\in\mathbb{N}} \subseteq \mathcal{M}^+$ can be used to construct an information complete POVM.

Proposition V.2.1. *Let $\mathcal{M}$ be a von Neumann algebra and $(\varphi_n)_{n\in\mathbb{N}} \subseteq \mathcal{M}_* \subset \mathcal{M}^*$ be a norm dense subset of $\mathcal{M}^*$. Suppose $(a_{i,j})_{i,j\in\mathbb{N}} \subseteq \mathcal{M}^+$ is constructed as in* (5.10)*. Let Ω be the collection of subsets of $\mathbb{N}\times\mathbb{N}$. Then define the function $\mu : \Omega \to \mathcal{M}^+$ as*

$$\mu(i,j) := a_{i,j}. \tag{5.12}$$

We can extend μ to Ω by linearity, namely for $A \subseteq \mathbb{N}\times\mathbb{N}$, let

$$\mu(A) := \sum_{(i,j)\in A} \mu(i,j). \tag{5.13}$$

Then μ is an IC-POVM.

Proof. We will first show μ is a POVM. Let $i,j \in \mathbb{N}$, then by construction $a_{i,j} \geq 0$, and $||a_{i,j}|| = p(i,j) < 1$. By the Spectral Mapping Theorem (B.0.5),

$$\sigma(\mathbf{1} - a_{i,j}) = 1 - \sigma(a_{i,j}) \subseteq [0,1]. \tag{5.14}$$

Hence for $i \neq j$, we have $0 \leq \mu(i,j) \leq \mathbf{1}$. Moreover,

$$\mu(\mathbb{N}\times\mathbb{N}) = \sum_{(i,j)\in\mathbb{N}\times\mathbb{N}} a_{i,j} = a_{0,0} + \sum_{i,j=0,i\neq j}^{\infty} a_{i,j} = \mathbf{1}. \tag{5.15}$$

Note that $0 \leq a_{0,0} \leq \mathbf{1}$ since,

$$\Big|\Big| \sum_{i,j=0,i\neq j} a_{i,j} \Big|\Big| \leq 1, \tag{5.16}$$

and $\sigma(\mathbf{1} - \sum_{i,j=0,i\neq j} a_{i,j}) \subseteq [0,1]$. Then μ is a positive operator valued measure since the strong operator topology is weaker than the norm topology. Now we will show that μ is information complete. The implication from left to right is trivial. We will show the other implication by a proof by contradiction. Let $\tau,\gamma \in \mathcal{M}_*$, and suppose

$$\forall i,j \in \mathbb{N}, \quad \tau(\mu(i,j)) = \gamma(\mu(i,j)), \tag{5.17}$$

while $\tau \neq \gamma$. Then there exists a $\delta > 0$, such that

$$||\tau - \gamma|| = \delta. \tag{5.18}$$

Let $r = \frac{1}{20}\cdot\delta$. Consider the two open balls $B_r(\tau) = \{\varphi \in \mathcal{M}_*, ||\varphi - \tau|| < r\}$ and $B_r(\gamma)$. Because $\mathcal{M}_*$ is separable we know there exists an $n,m \in \mathbb{N}$ such that $\varphi_n \in B_r(\tau)$ and $\varphi_m \in B_r(\gamma)$. Since $||\tau - \gamma|| = \delta$ we also know,

$$(\delta - 2r) = \frac{9}{10}\delta \leq ||\varphi_n - \varphi_m||. \tag{5.19}$$

Recall, $a_{n,m} = p(n,m)\tilde{a}_{n,m}$ for $n \neq m$, and by linearity,

$$\tau(\tilde{a}_{n,m}) = \gamma(\tilde{a}_{n,m}). \tag{5.20}$$

Moreover, by construction of the elements $(\tilde{a}_{i,j})_{i,j\in\mathbb{N}}$ in Lemma V.2.1,

$$\frac{1}{5}||\varphi_n - \varphi_m|| \leq |\varphi_n(\tilde{a}_{n,m}) - \varphi_m(\tilde{a}_{n,m})|. \tag{5.21}$$

Combining these observations we have

$$\begin{aligned}\frac{1}{5}||\varphi_n - \varphi_m|| \leq & |\varphi_n(\tilde{a}_{n,m}) - \varphi_m(\tilde{a}_{n,m})| \\ = & |\tau(\tilde{a}_{n,m}) - \gamma(\tilde{a}_{n,m}) - \varphi_n(\tilde{a}_{n,m}) + \varphi_m(\tilde{a}_{n,m})| \\ \leq & |\tau(\tilde{a}_{n,m}) - \varphi_n(\tilde{a}_{n,m})| + |\gamma(\tilde{a}_{n,m}) - \varphi_m(\tilde{a}_{n,m})| \\ \leq & 2r = \frac{1}{10}\delta. \end{aligned} \tag{5.22}$$

But then we get the contradiction by recalling (5.19),

$$\Big(\frac{9}{10}\delta\Big)\frac{1}{5} \leq \frac{1}{5}||\varphi_n - \varphi_m|| \leq \frac{1}{10}\delta \implies \frac{9}{10} \leq \frac{1}{2} \quad \lightning, \tag{5.23}$$

and $\tau = \gamma$. □

V.2.2 Constructed from a norm separable proper ideal

In this section we present a construction of a IC-POVM which is very similar to what we did in Proposition V.2.1. However, as opposed to using the collection of operators $(\tilde{a}_{i,j})_{i,j\in\mathbb{N}}$ from Lemma V.2.1, we will use a countable norm dense subset of a proper ideal (Definition B.0.4) in a factor $\mathcal{M}$. It will turn out that the IC-POVM we construct here is fundamentally different than the one constructed in the previous section. The motivation for our construction is the following two propositions. Note, all ideals we refer to are two-sided norm-closed ideals.

Proposition V.2.2. *If $\mathcal{M}$ is a factor, then every non-zero ideal $J \subseteq \mathcal{M}$ is $\sigma-$weakly dense in $\mathcal{M}$.*

Proof. See [35, Proposition 3.12]. □

Proposition V.2.3. *Let $\mathcal{M}$ be a von Neumann algebra and $J \subset \mathcal{M}$ a proper ideal. Then, if $\psi : J \to \mathbb{C}$ is continuous and positive linear functional on J, there exists a unique extension of ψ, to a positive linear functional ϕ on $\mathcal{M}$ such that $||\psi|| = ||\phi||$.*

Proof. See [7, Proposition II.6.4.16]. □

The key example confirming proper norm-closed ideals exist for type I_∞ von Neumann algebras is the following proposition:

Proposition V.2.4. *If $\mathcal{H}$ is a infinite dimensional separable Hilbert space, then the set of compact operators $K(\mathcal{H}) \subset B(\mathcal{H})$ (Definition B.0.2), is a proper ideal of $B(\mathcal{H})$. Moreover, $K(\mathcal{H})$ is norm separable, unlike $B(\mathcal{H})$.*

Proof. See [7, I.8.1.2]. □

Now, we can construct an IC-POVM in the case that $\mathcal{M}$ is a factor with a proper norm-closed, and norm-separable, ideal $J \subset \mathcal{M}$.

Proposition V.2.5. *Let $\mathcal{M}$ be a factor with a proper norm-separable ideal $J \subset \mathcal{M}$.*

1. *Then there exists a subset $(\tilde{a}_n)_{n\in\mathbb{N}} \subseteq \mathcal{M}^+ \cap J$, such that for all $n \in \mathbb{N}$,*

$$\mathrm{span}_{\mathbb{C}}\{\tilde{a}_n; n \in \mathbb{N}\}^- = J, \tag{5.24}$$

   where the closure is taken with respect to the norm topology on $\mathcal{M}$.

2. *Fix $p \in \mathrm{Prob}(\mathbb{N})$, such that for all $n \in \mathbb{N}$, $p(n) > 0$. Then for $n \in \mathbb{N}$, define $a_n \in J$ as*

$$\begin{aligned} a_n &:= \frac{p(n)}{||\tilde{a}_n||}\tilde{a}_n, \quad n > 0, \\ a_0 &:= \boldsymbol{1} - \sum_{n=1}^{\infty} a_n. \end{aligned} \tag{5.25}$$

Let Ω be the collection of subsets of $\mathbb{N}$. Then the function $\mu : \Omega \to J \cap \mathcal{M}^+$ defined pointwise as

$$\mu(n) := a_n, \tag{5.26}$$

which we extended by linearity to subsets $A \subseteq \Omega$, is an IC-POVM for $\mathcal{M}$.

Proof. 1. Since J is norm separable, let $(b_n)_{n\in\mathbb{N}} \subseteq J$ be a norm dense subset. Moreover, $J \subset \mathcal{M} \subseteq B(H)$ is a C^* subalgebra of $B(H)$. Therefore, for every $n \in \mathbb{N}$, we use the polar identity to decompose b_n, into a linear combination of 4 positive operators, namely:

$$b_n = u_1 - u_2 + i(v_1 - v_2), \quad \text{where } u_1, u_2, v_1, v_2 \in J \cap \mathcal{M}^+. \tag{5.27}$$

Note, an operator $u \in J$ is positive iff for all $|\psi\rangle \in \mathcal{H}$, we have $\langle\psi, u\psi\rangle \geq 0$, which implies $u \in J \cap \mathcal{M}^+$. Using (5.27), for all $n \in \mathbb{N}$, let $\tilde{a}_{n_1}, \tilde{a}_{n_2}, \tilde{a}_{n_3}, \tilde{a}_{n_4} \in J \cap \mathcal{M}^+$, such that $b_n = \tilde{a}_{n_1} - \tilde{a}_{n_1} + i(\tilde{a}_{n_3} - \tilde{a}_{n_4})$. Then since $(b_n)_{n\in\mathbb{N}}$ is norm-dense in J, we know that

$$\mathrm{span}_{\mathbb{C}}\{\tilde{a}_n; n \in \mathbb{N}\}^- = J. \tag{5.28}$$

2. We refer to reader to Proposition V.2.1, where we prove a *nearly identical* function μ is a POVM. What is left to show is that it is an IC-POVM. Let $\tau, \gamma \in \mathcal{M}_*$, and suppose

$$\forall n \in \mathbb{N}, \quad \tau(\mu(n)) = \gamma(\mu(n)). \tag{5.29}$$

Then, by part *1*, we know for all $a \in J$, $\gamma(a) = \tau(a)$. This is because for all linear combinations $u \in \mathrm{span}_{\mathbb{C}}\{\tilde{a}_n; n \in \mathbb{N}\}$, $\gamma(u) = \tau(u)$, and because γ, τ are both σ-weakly continuous they are also norm continuous. Therefore, for all sequence $(u_n)_{n\in\mathbb{N}} \subseteq \mathrm{span}_{\mathbb{C}}\{\tilde{a}_n; n \in \mathbb{N}\}$, converging in norm to some $u \in J$, we get $\lim_{n\to\infty} \gamma(u_n) = \lim_{n\to\infty} \tau(u_n) = \gamma(u) = \tau(u)$. By Proposition V.2.2, we have $J^{-\sigma w} = \mathcal{M}$, and similarly, for all nets $(a_i)_{i\in I} \subseteq J$ converging σ-weakly to $a \in \mathcal{M}$, we have

$$\lim_{i\in I} \gamma(a_i) = \lim_{i\in I} \tau(a_i) = \gamma(a) = \tau(a), \tag{5.30}$$

and $\tau = \gamma$. □

A natural question is; which von Neumann algebras have a proper norm-separable ideal? While slightly disappointing, the answer is not many. Recall, a C^* algebra is called **simple** if it contains no non-trivial two sided closed ideals. We then have the following remarkable classifications of the ideals of factors.

Proposition V.2.6. *Let $\mathcal{M} \subseteq B(\mathcal{H})$ be a factor and $\mathcal{H}$ be a separable Hilbert space.*

1. *If $\mathcal{M}$ is type I_n or type II_1 then $\mathcal{M}$ is a simple.*
2. *If $\mathcal{M}$ is type III, then $\mathcal{M}$ is simple.*
3. *If $\mathcal{M}$ is type I_∞ or type II_∞, then $\mathcal{M}$ has a unique non-trivial norm-closed ideal K.*

Proof. See [7, III.1.7.11]. □

By Proposition V.2.6, we see that the situations where the POVM constructed in Proposition V.2.5, could be useful is if $\mathcal{M}$ is either type I_∞ or type II_∞. Unfortunately, it turns out that the proper ideal for a type II_∞ factor, need not norm-separable. Thus the POVM μ we have constructed in this section is only suitable for when $\mathcal{M} = B(\mathcal{H})$ is a type I_∞ von Neumann algebra, with the proper norm-separable ideal $J = K(\mathcal{H})$. In the next section, we will see the POVM constructed in Proposition V.2.1 is an insufficient tool for extending the quantum Sanov theorem, and how the construction in Proposition V.2.5 provides a remedy for when $\mathcal{M} = B(\mathcal{H})$.

V.3 Extending the universal Sanov theorem

V.3.1 Type I_∞

In this section we extend the universal quantum Sanov theorem from the setting of a type I_d von Neumann algebra to a von Neumann algebra of type I_∞.

Definition V.3.1 (Weak-$*$ topology on $\mathcal{S}_n(\mathcal{M})$)**.** *The **weak-$*$** topology $\mathcal{S}_n(\mathcal{M}) \subseteq \mathcal{M}_*$, is the weak-$*$ topology on $\mathcal{M}_*$. Namely, using that the Banach dual of $\mathcal{M}_*$ is given by $(\mathcal{M}_*)^* \simeq \mathcal{M}$ (Theorem V.1.1), we have that a net of normal states $(\psi_i)_{i\in I} \subseteq \mathcal{S}_n(\mathcal{M})$, converges to a normal state $\psi \in \mathcal{S}_n(\mathcal{M})$ $*$-weakly, iff for all $a \in \mathcal{M}$*

$$\lim_{i\in I} \varphi_i(a) = \varphi(a). \tag{5.31}$$

Proposition V.3.1. *Let $\mathcal{M} = B(\mathcal{H})$, where $\mathcal{H}$ is separable, be a type I_∞ von Neumann algebra. Let $J = K(\mathcal{H}) \subset \mathcal{M}$ be the unique two-sided norm-closed proper ideal in $\mathcal{M}$. Let $X = \mathbb{N}$ and Ω the collection of all subsets of X. Take $\mu : \Omega \to \mathcal{M}^+$ to be a IC-POVM as constructed in Proposition V.2.5. Define*

$$\Lambda := \{p \in Prob(X); \exists \varphi \in \mathcal{S}_n(\mathcal{M}) \text{ s.t } \forall A \in \Omega, p(A) = \varphi(\mu(\Omega))\}, \tag{5.32}$$

then there is a function $\Phi : \Lambda \to \mathcal{S}_n(\mathcal{M})$ that is bijective and continuous with respect to the weak topology on $\Lambda \subseteq Prob(X)$ and the weak-$*$ topology on $\mathcal{S}_n(\mathcal{M}) \subseteq \mathcal{M}_*$. We call Φ a complete tomography.

Proof. Because μ is an IC-POVM, for each $p \in \Lambda$ there exists only one $\varphi \in \mathcal{S}_n(\mathcal{M})$, such that for all measurable $A \subseteq \Omega$

$$p(A) = \varphi(\mu(A)). \tag{5.33}$$

Define $\Phi : \Lambda \to \mathcal{S}_n(\mathcal{M})$ as,

$$\Phi(p) := \varphi, \tag{5.34}$$

where φ is the unique normal state satisfying $\forall A \in \Omega, p(A) = \varphi(\mu(A))$. Clearly Φ is an bijective map. Now we will show Φ is continuous with respect to the weak topology on Λ and the weak-$*$ topology on $\mathcal{S}_n(\mathcal{M})$. Let $(p_i)_{i\in I} \subseteq \Lambda$ be a net of operators converging weakly to $p \in \Lambda$. Since $X = \mathbb{N}$ is Polish, the weak topology on $\mathrm{Prob}(X)$ is generated by the Lévy metric Theorem II.1.1, and $\mathrm{Prob}(X)$ is Polish. We therefore assume the net $(p_n)_{n\in\mathbb{N}}$ is a sequence. The aim of the rest of the proof is to show that $\Phi(p_n)$ conveges $*$-weakly to $\Phi(p)$ as $n \to \infty$, or, for all $a \in \mathcal{M}$

$$\lim_{n\to\infty} \Phi(p_n)(a) = \Phi(p)(a). \tag{5.35}$$

For the rest of this proof we use the following notation. For all $n \in \mathbb{N}$, let

$$\varphi_n := \Phi(p_n), \quad \varphi := \Phi(p). \tag{5.36}$$

Because $\varphi \in \mathcal{S}_n(\mathcal{M})$ and $\varphi_n \in \mathcal{S}_n(\mathcal{M})$ are states, we know $||\varphi|| = 1$ and $||\varphi_n|| = 1$ for all $n \in \mathbb{N}$. Moreover, by Banach-Alaoglu's Theorem B.0.1, the closed unit ball

$$B^* := \{\gamma \in \mathcal{M}^*; ||\gamma|| \leq 1\}, \tag{5.37}$$

is weak-$*$ compact in $\mathcal{M}^*$. Therefore, by Proposition A.1.4, we know the sequence $(\varphi_n)_{n\in\mathbb{N}} \subset B^*$, has a subnet (Definition A.1.4) $(\varphi_j)_{j\in J}$, which $*$-weakly converges to some $\varphi' \in \mathcal{M}^*$. That is for all $a \in \mathcal{M}$, we have

$$\lim_{j\in J} \varphi_j(a) = \varphi'(a). \tag{5.38}$$

To show (5.35), we in fact only need to prove that $\varphi' = \varphi$. This is due to the following. If $\varphi' = \varphi$, then any converging subnet $(\varphi_j)_{j\in J}$ of $(\varphi_n)_{n\in\mathbb{N}}$, necessarily converges to φ. Hence, by Proposition A.1.2, φ is a unique accumulation point (Definition A.1.3) of $(\varphi_n)_{n\in\mathbb{N}}$. Finally, since B^* is compact and the weak-$*$ topology is Hausdorff, by Proposition A.1.5, $(\varphi_n)_{n\in\mathbb{N}}$ converges $*$-weakly to φ, and (5.35) holds. Now we will prove $\varphi' = \varphi$.

First, by definition $\{\varphi_j; j \in J\} \subseteq \mathcal{S}_n(\mathcal{M})$, therefore for all positive operators $a \in \mathcal{M}^+$ we have

$$\lim_{j\in J} \varphi_j(a) = \varphi'(a) \geq 0, \tag{5.39}$$

and

$$\lim_{j\in J} \varphi_j(\mathbf{1}) = \varphi'(\mathbf{1}) = 1. \tag{5.40}$$

Hence $\varphi' \in \mathcal{S}(\mathcal{M})$ is a state. By construction of μ, we know by Proposition V.2.5 that

$$\mathrm{span}_{\mathbb{C}}\{\mu(n); n \in \mathbb{N}\}^- = K(\mathcal{H}). \tag{5.41}$$

Let $u \in K(\mathcal{H})$ be a compact operator. Take an arbitrary sequence

$$(u_m)_{m\in\mathbb{N}} \subset \mathrm{span}_{\mathbb{C}}\{\mu(n); n \in \mathbb{N}\}, \tag{5.42}$$

such that $\lim_{m\to} u_m = u$ in norm. By assumption we know for all $m \in \mathbb{N}$,

$$\lim_{n\to\infty} \varphi_n(u_m) = \varphi(u_m), \tag{5.43}$$

hence

$$\lim_{j\in J}\varphi_j(u_m)=\varphi'(u_m)=\varphi(u_m). \tag{5.44}$$

But since, $\varphi'\in\mathcal{S}(\mathcal{M})$ and $\varphi\in\mathcal{S}_n(\mathcal{M})$ are both norm-continuous, we know

$$\lim_{m\to\infty}\varphi'(u_m)=\varphi'(u)=\varphi(u)=\lim_{m\to\infty}\varphi(u_m). \tag{5.45}$$

And on $J=K(\mathcal{H})$ the two states φ' and φ are equal. Finally, by Proposition V.2.3, since $\varphi':J\to\mathbb{C}$ is a positive linear functional on the proper ideal J, it has a unique extension to a positive linear ϕ on $\mathcal{M}=B(\mathcal{H})$ such that $||\varphi'||=||\phi||$. But, φ is a positive linear functional on $\mathcal{M}$, and for all $u\in J$ we have $\varphi'(u)=\phi(u)=\varphi(u)$. Since $||\varphi'||=||\varphi||=1$, the unique extension ϕ of ϕ' is φ, and for all $a\in\mathcal{M}$, we have $\varphi'(a)=\phi(a)=\varphi(a)$, concluding the proof. □

In the setting of Proposition V.3.1, let $n\in\mathbb{N}$ and $L_n:X^n\to\text{Prob}(X)$ be the empirical measure. Note that X is a Polish space, which will be used later when we apply the classical Sanov theorem.

Consider the n-fold tensor product von Neumann algebra $\mathcal{M}^{\otimes n}$. We can extend any normal state $\varphi\in\mathcal{S}_n(\mathcal{M})$ to a normal state $\varphi^{\otimes n}\in\mathcal{S}_n(\mathcal{M}^{\otimes n})$. Similar to extending a state onto the tensor product of $\mathcal{M}^{\otimes n}$, we can extend the POVM $\mu:\Omega\to\mathcal{M}^+$ to a POVM $\mu_n:\Omega^n\to(\mathcal{M}^{\otimes n})^+$ for all $n\in\mathbb{N}$. Namely, let $A\subseteq\Omega^n$ and define

$$\mu_n(A):=\sum_{a\in A}\mu(a_1)\otimes\cdots\otimes\mu(a_n). \tag{5.46}$$

Note by Ω^n we denote the n-fold product σ-algebra of Ω and, by Proposition V.1.1 we know μ_n is indeed a POVM. Finally, we can state the extension of the universal quantum Sanov theorem.

Theorem V.3.1. *Suppose $\mathcal{H}$ is a separable Hilbert space. Let $\mathcal{M}=B(\mathcal{H})$ be a von Neumann algebra, and let $\mu:\Omega\to B(\mathcal{H})$ be the IC-POVM as constructed in Proposition V.2.5. Let $\mathcal{A}\subseteq S_n(\mathcal{M})$ be $*$-weakly measurable set. For all $n\in\mathbb{N}$, define the test $T_n=\mu_n(L_n^{-1}(\Phi^{-1}(\mathcal{A})))$. Then, for all $\varphi\in\mathcal{S}_n(\mathcal{M})$,*

$$\begin{aligned}-\inf_{\gamma\in\mathcal{A}^\circ}D_{KL}(\Phi^{-1}(\gamma)||\Phi^{-1}(\varphi))&\le\liminf_{n\to\infty}\frac{1}{n}\log\varphi^{\otimes n}(T_n)\\&\le\limsup_{n\to\infty}\frac{1}{n}\log\varphi^{\otimes n}(T_n)\\&\le-\inf_{\gamma\in\mathcal{A}^-}D_{KL}(\Phi^{-1}(\gamma)||\Phi^{-1}(\varphi)),\end{aligned} \tag{5.47}$$

where $\mathcal{A}^\circ,\mathcal{A}^-$ are respectively the weak-$*$ interior and closure of $\mathcal{A}$.

Proof. Let $G=\Phi^{-1}(\mathcal{A})\subseteq\Lambda\subseteq\text{Prob}(\Omega)$. By Proposition V.3.1 Φ is weakly and $*$-weakly continuous, and since $\mathcal{A}$ is $*$-weakly measurable, we know G is weakly measurable. Let $q=\Phi^{-1}(\varphi)$. By Theorem II.3.1,

$$\begin{aligned}-\inf_{p\in G^\circ}D_{KL}(p||q)&\le\liminf_{n\to\infty}\frac{1}{n}\log q^n(L_n\in G)\\&\le\limsup_{n\to\infty}\frac{1}{n}\log q^n(L_n\in G)\le-\inf_{p\in G^-}D_{KL}(p||q).\end{aligned} \tag{5.48}$$

By construction we know

$$q^n(L_n \in G) = q^n(L_n^{-1} \circ \Phi^{-1}(\mathcal{A})) = \varphi^{\otimes n}(T_n). \tag{5.49}$$

Because Φ is a bijection for any $\mathcal{Q} \subseteq \mathcal{S}_n(\mathcal{M})$,

$$\inf_{\gamma \in \mathcal{Q}} D(\Phi^{-1}(\gamma)||\Phi^{-1}(\phi)) = \inf_{p \in \Phi^{-1}(\mathcal{Q})} D(p||q), \tag{5.50}$$

concluding the proof. □

V.3.2 A counter example constructed by Vaes

In this chapter we constructed two possible IC-POVM's for infinite dimensional von Neumann algebras: one being Proposition V.2.5, which turned out to only be a suitable construction for type I_∞ von Neumann algebras, and the other being Proposition V.2.1, which was suitable for any von Neumann algebra $\mathcal{M}$ with a norm-separable predual. A natural question is, can one make use of the IC-POVM defined in Proposition V.2.1 to extend our quantum Sanov theorem to a more general setting? The short answer is, we were not able to successfully do this.

The main difficulty we faced can be summarized as follows. Let $\mu' : \Omega \to \mathcal{M}^+$ be the IC-POVM constructed as in Proposition V.2.1 and let (X, Ω) be the underlying measure space. Any attempt of extending our universal quantum Sanov theorem relies on generalizing Proposition V.3.1; where we need to construct a complete tomography $\Phi' : \Lambda' \to \mathcal{S}_n(\mathcal{M})$ for some convex non-empty set $\Lambda' \subset \mathrm{Prob}(X)$. Define $\Lambda' \subset \mathrm{Prob}(X)$ as in (5.32) and $\Phi' : \Lambda' \to \mathcal{S}_n(\mathcal{M})$ as in (5.34) with respect to μ'. By definition of an IC-POVM it follows that Φ' is bijective, and the only difficulty is to prove Φ' is continuous with respect some topology on $\Lambda' \subset \mathrm{Prob}(X)$ and a topology on $\mathcal{S}_n(\mathcal{M})$. Since $X = \mathbb{N}\times\mathbb{N}$, the weak and τ-topology coincide on Λ', so assume $(p_n)_{n\in\mathbb{N}} \subseteq \Lambda'$ is some sequence that converges weakly to $p \in \Lambda'$. For all $n \in \mathbb{N}$, define $\gamma_n \in \mathcal{S}_n(\mathcal{M})$ and $\gamma \in \mathcal{S}_n(\mathcal{M})$ as,

$$\gamma_n := \Phi'(p_n), \qquad \gamma := \Phi'(p). \tag{5.51}$$

Now define the set,

$$B_0 := \mathrm{span}_{\mathbb{C}}\{\mu'(i,j); (i,j) \in X\}, \tag{5.52}$$

then by Proposition V.1.2, $B_0^{\sigma-w} = \mathcal{M}$, and we know for all $b \in B_0$,

$$\lim_{n\to\infty} \gamma_n(b) = \gamma(b). \tag{5.53}$$

Now let $B = B_0^-$ be the norm closure of B_0. Assume $a \in B$ and let $(b_m)_{m\in\mathbb{N}} \subset B_0$ be a sequence converging to a. Then for all $n, m \in \mathbb{N}$,

$$\begin{aligned} |\gamma_n(a) - \gamma(a)| &= |\gamma_n(a) - \gamma_n(b_m) + \gamma_n(b_m) - \gamma(a) + \gamma(b_m) - \gamma(b_m)| \\ &\leq |\gamma_n(a) - \gamma_n(b_m)| + |\gamma(a) - \gamma(b_m)| + |\gamma_n(b_m) - \gamma(b_m)|. \end{aligned} \tag{5.54}$$

Let $\varepsilon > 0$, and fix an $m \in \mathbb{N}$ such that $||a - b_m|| < \frac{\varepsilon}{3}$. Then since γ and γ_n are continuous linear states for all $n \in \mathbb{N}$,

$$|\gamma_n(a) - \gamma_n(b_m)| + |\gamma(a) - \gamma(b_m)| \leq \frac{2\varepsilon}{3}. \tag{5.55}$$

Since $\lim_{n\to\infty}\gamma_n(b_m)=\gamma(b_m)$, we then know there is an $N\in\mathbb{N}$ such that for all $n\geq N$,

$$|\gamma_n(a)-\gamma(a)|<\frac{\varepsilon}{3}, \tag{5.56}$$

and we can conclude $\lim_{n\to\infty}\gamma_n(a)=\gamma(a)$. To summarize, we are faced with the following question. Let B be a norm closed subset of $\mathcal{M}$, and assume B is σ-weakly dense in $\mathcal{M}$. If $\gamma,\gamma_n\in\mathcal{M}_*$ are normal states for all $n\in\mathbb{N}$, and for all $a\in B$, $\gamma_n(a)$ converges to $\gamma(a)$, then does $\gamma_n(a)$ converges to $\gamma(a)$ for all $a\in\mathcal{M}$?

The answer to this question is no. Before we provide a counter example, we briefly discuss the implications of this result. We believe that most likely, the IC-POVM constructed in Proposition V.2.1 is an insufficient tool for constructing a complete tomography as in Proposition V.3.1. The fact that we succeed in using Proposition V.2.5 to construct an complete tomography, is largely due to Proposition V.2.3, which allowed us to use the special properties of proper ideals in von Neumann algebras. However, Proposition V.2.6 guarantees that the POVM constructed in Proposition V.2.5 cannot be used beyond the type I_∞ case, since it is the only setting where a norm-separable closed ideal exists in a von Neumann algebra. These results hint that likely, a more subtle approach will be necessary if one wishes to obtain a universal quantum Sanov theorem for type II and type III factors.

As promised, we provide the following counter example to our most recent question.[2] Let $Y=\{0,1\}^{\mathbb{N}}$ and equip Y with the cylinder σ-algebra. Let $f\in\mathrm{Prob}(\{0,1\})$ be the flat distribution. By the Carathéodory extension theorem we can define the probability distribution $p=f^{\mathbb{N}}\in\mathrm{Prob}(Y)$. Define $\mathcal{M}=L^2(Y,p)$, and let $\gamma:L^\infty(Y,p)\to\mathbb{C}$ be the normal state defined by p;

$$\gamma(h):=\int_Y h(x)\mathrm{d}p(x). \tag{5.57}$$

For all $n\in\mathbb{N}$, define the C^* sub-algebra $B_n\subset\mathcal{M}$ as

$$B_n:=\{f\in\mathcal{M};f\text{ depends only on the first }n\text{-coordinates of }x\in Y\}. \tag{5.58}$$

Define the C^* sub-algebra $B\subseteq\mathcal{M}$,

$$B:=\Big(\bigcup_{n\in\mathbb{N}}B_n\Big)^-, \tag{5.59}$$

which is σ-weakly dense in $\mathcal{M}$. We now construct a sequence of normal states $\gamma_n\in\mathcal{M}_*$, which converges pointwise for all $b\in B$ to the normal state $\gamma\in\mathcal{M}_*$, while there exists a projection $P\in\mathcal{M}$ such that $\gamma_n(P)=1$ for all $n\in\mathbb{N}$, and $\gamma(P)<1$. For $k\in\mathbb{N}$, let $U_k\subseteq Y$ and $U\subseteq Y$ be defined as:

$$\begin{aligned}U_k&:=\{x\in Y;x_k=x_{k+1}=\cdots=x_{2k}=0\},\\ U&:=\bigcup_{k\in\mathbb{N}}U_k.\end{aligned} \tag{5.60}$$

Then for all $k\in\mathbb{N}$, $p(U_k)=\frac{1}{2^{k+1}}$, and $p(U)\leq\frac{1}{2}$. Let $P=\mathbb{I}_U$, be the indicator function for $U\subset Y$, then by construction $P\in\mathcal{M}$ is a projection and $\gamma(P)\leq\frac{1}{2}$.

[2]This counter example was kindly provided by Stefaan Vaes.

Now we will construct the sequence of normal states γ_n. For all $n \in \mathbb{N}$, we have that B_n is a finite dimensional C^* subalgebra of $\mathcal{M}$. Consider the $*$-homomorphism

$$\begin{aligned} \pi_n : B_n &\longrightarrow \mathcal{M}, \\ f &\longmapsto f \cdot \mathbb{I}_{U_n}. \end{aligned} \tag{5.61}$$

Since f is only dependent on the first n-coordinates of $x \in Y$, and U_n is only dependent on the coordinates $\{x_n, \cdots, x_{2n}\}$ in $x \in Y$, we know $\ker \pi_n = \{0\}$ and π_n is faithful. Similarly, the $*$-homomorphism defined as

$$\begin{aligned} \pi : B &\longrightarrow \mathcal{M}, \\ f &\longmapsto f \cdot P, \end{aligned} \tag{5.62}$$

is also faithful. We can therefore define the following states. Define the finite dimensional C^* subalgebra $C_n := B_n P + B_n(\mathbf{1} - P)$ and for $(aP + b(\mathbf{1} - P)) \in C_n$, define the state

$$\omega_n(bP + c(\mathbf{1} - P)) := \gamma(b). \tag{5.63}$$

Then for all $b \in B_n$, we have $\omega_n(bP) = \omega_n(b) = \gamma(b)$ and $\omega_n(P) = \omega_n(\mathbf{1}P) = \gamma(\mathbf{1}) = 1$. Since C_n is a finite dimensional C^* subalgebra of $\mathcal{M}$, we can extend ω_n to a normal state $\gamma_n : \mathcal{M} \to \mathbb{C}$, such that for all $k \in \mathbb{N}$ and $b \in B_k$ if $n \geq k$, then

$$\omega_n(b) = \gamma_n(b) = \gamma(b). \tag{5.64}$$

Therefore for all $b \in \cup_{k \in \mathbb{N}} B_k$ the sequence $\gamma_n(b)$ converges to $\gamma(b)$. Using the same argument as in (5.56), we can conclude for all $b \in B$, $\lim_{n \to \infty} \gamma_n(b) = \gamma(b)$, while

$$\gamma_n(P) = \omega_n(P) = 1, \tag{5.65}$$

for all $n \in \mathbb{N}$ and $\gamma(P) \leq \frac{1}{2}$. Hence, γ_n does not converge pointwise for all $b \in \mathcal{M}$ to γ while it converges pointwise to γ for all $b \in B$.

Chapter VI

Conclusion and outlook

In this work we explored several approaches to extend the classical Sanov theorem to the quantum regime. In Chapter II we showcased how the classical Sanov theorem can be used to prove several foundational results in probability theory, such as the law of large numbers Proposition II.2.1, Stein's lemma Lemma II.2.2 and Chernoff's lemma Lemma II.2.3. We believe the classical Sanov theorem describes the interaction we as observes have with (classical) randomness. This viewpoint encouraged us to establish a Sanov theorem in the quantum regime.

In Chapter III we presented two different approaches from literature to extend the classical Sanov theorem to the quantum regime. We concluded, the first proposed quantum Sanov theorem in [6], is a quantum extension of classical result Theorem II.2.1. Additionally, the result proposed by [16] (Theorem III.5.1) is closely related to our own research, however, it is only concerned with universally separating a fixed state $\rho \in D(\mathcal{H})$ from an arbitrary state $\sigma \in D(\mathcal{H})$.

Motivated by the two previously quantum Sanov theorem's, we proposed our own universal quantum Sanov theorem in Theorem IV.2.1. In this result we constructed a quantum empirical measure $\hat{L}_n^\mu : \Omega^n \to D(\mathcal{H})$, which is induced by an IC-POVM $\mu : \Omega \to D(\mathcal{H})$. The main feature of our result is the universality property of the sequence of separating tests T_n we constructed. Additionally, in Theorem IV.1 we prove the rate function $D_\Phi : D(\mathcal{H}) \times D(\mathcal{H}) \to [0, \infty]$ we obtain in Theorem IV.2.1, is necessarily bounded above by the Umegaki relative entropy $D : D(\mathcal{H}) \times D(\mathcal{H}) \to [0, \infty]$, since universal state separation is not possible at a rate given by D. Theorem IV.1 establishes an interesting difference between classical and quantum randomness, since the existence of non-commuting quantum states $\sigma, \rho \in D(\mathcal{H})$ was essential for our proof, and in classical probability all distributions commute (upon embedding them into a quantum system). In the setting of Chapter IV, possible directions for future research can be motivated by the following questions:

- What is the optimal IC-POVM μ to use when constructing a sequence of universally separating tests as in Lemma IV.2.1?

It is clear that different IC-POVM's μ_1, μ_2 can achieve different separation rates, i.e. for specific $\sigma, \rho \in D(\mathcal{H})$,

$$D_{\Phi_1}(\rho||\sigma) < D_{\Phi_2}(\rho||\sigma), \tag{6.1}$$

where Φ_i is the induced complete tomography of μ_i for $i = 1, 2$. Thus optimality can be viewed as maximizing the entropy D_Φ. Moreover, another interesting question is:

- When are two IC-POVM's equivalent in the sense that $D_{\Phi_1}(\rho||\sigma) = D_{\Phi_2}(\rho||\sigma)$ for all $\sigma, \rho \in D(\mathcal{H})$?

In Chapter V, we used von Neumann algebras to extend Theorem IV.2.1 to type I_∞ factors. We also presented the difficulties we faced when attempting to extend IV.2.1 to factors of type II and III. Given more time, it would be interesting to establish a universal quantum Sanov theorem for arbitrary von Neumann algebra $\mathcal{M}$. A possible direction would be to make use of the Sanov theorem for general measure spaces described in Section II.3. Namely:

- Does their exist a measure space (X, Ω) and an IC-POVM $\mu : \Omega \to \mathcal{M}$ such that if

$$\Lambda := \{p \in \mathrm{Prob}(X); \exists \varphi \in \mathcal{S}_n(\mathcal{M}) \text{ such that } \forall A \in \Omega, p(A) = \varphi(\mu(A))\}, \qquad (6.2)$$

  then the bijective map $\Phi : \Lambda \to \mathcal{S}_n(\mathcal{M})$ (as constructed in Proposition IV.2.1) is continuous with respect to the τ-topology on Λ and the weak-$*$ topology on $\mathcal{S}_n(\mathcal{M})$?

Besides for the possible research directions directly related to this thesis, we view this work as the initial steps needed to extend more results from classical large deviation theory to the quantum regime.

Bibliography

[1] A. D. Acosta. On large deviations of empirical measures in the tau-topology. *Journal of Applied Probability*, 31:41–47, 1994. ISSN 0021-9002. doi: 10.2307/3214947.

[2] K. B. Athreya and S. N. Lahiri. *Measure Theory and Probability Theory*. Springer, 2006. ISBN 0387354344.

[3] G. Aubrun. A naive look at Schur–Weyl duality. URL `https://math.univ-lyon1.fr/~aubrun/recherche/schur-weyl.pdf`.

[4] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete. Discriminating states: The quantum chernoff bound. *Physical Review Letters*, 98(16), 2007. ISSN 1079-7114. URL `http://dx.doi.org/10.1103/PhysRevLett.98.160501`.

[5] I. Bjelakovi´c and R. Siegmund-Schultze. Quantum stein's lemma revisited, inequalities for quantum entropies, and a concavity theorem of lieb. Technical report, 2012.

[6] I. Bjelaković, J. D. Deuschel, T. Krüger, R. Seiler, R. Siegmund-Schultze, and A. Szkoła. A Quantum Version of Sanov's Theorem. *Communications in Mathematical Physics*, pages 659–671, 12 . ISSN 00103616. doi: 10.1007/s00220-005-1426-2.

[7] B. Blackadar. *Operator Algebras: Theory of C* Algebras and von Neumann Algebras*. Number v. 13 in Encyclopaedia of Mathematical Sciences. Springer, 2006. ISBN 9783540284864. URL `https://books.google.nl/books?id=6b_T1j3Ib8oC`.

[8] F. G. S. L. Brandão, A. W. Harrow, J. R. Lee, and Y. Peres. Adversarial hypothesis testing and a quantum stein's lemma for restricted measurements. Technical report. URL `https://doi.org/10.1145/2554797.2554816`.

[9] P. Busch, P. Lahti, J.-P. Pellonpaa, and K. Ylinen. *Quantum measurement*. Springer, 2016. ISBN 9783319433899.

[10] C. L. Canonne. A short note on an inequality between KL and TV, 2023. URL `https://arxiv.org/pdf/2202.07198`.

[11] J. Conway. *A Course in Functional Analysis*. Graduate Texts in Mathematics. Springer New York, 1994. ISBN 9780387972459. URL `https://books.google.nl/books?id=ix4P1e6AkeIC`.

[12] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2009. ISBN 9783642033117. URL `https://books.google.nl/books?id=iT9JRlGPx5gC`.

[13] N. Etemadi. An elementary proof of the strong law of large numbers. Technical report, 1981.

[14] P. I. Etingof. *Introduction to representation theory*. American Mathematical Society, 2011. ISBN 9780821853511.

[15] A. Grootveld, H. Yang, B. Chen, V. Gandikota, and J. Pollack. Towards quantum universal hypothesis testing. 7 2025.

[16] M. Hayashi. Another quantum version of Sanov theorem. 7 2024. URL http://arxiv.org/abs/2407.18566.

[17] M. Hayashi. General detectability measure, 2025. URL https://arxiv.org/abs/2501.09303.

[18] M. Hayashi and H. Yamasaki. Generalized quantum stein's lemma and second law of quantum resource theories, 2025. URL https://arxiv.org/abs/2408.02722.

[19] F. Hiai. *Quantum f-Divergences in von Neumann Algebras: Reversibility of Quantum Operations*. Mathematical Physics Studies. Springer Nature Singapore, 2021. ISBN 9789813341999. URL https://books.google.nl/books?id=41QXEAAAQBAJ.

[20] F. Hiai and D. Petz. Communications in Mathematical Physics The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability. Technical report, 1991.

[21] J. M. Hot, J. Pečarić, and V. Šimić. Inequalities involving the arithmetic and geometric operator means. *Mathematical Inequalities & Applications*, pages 415–430, 2008. ISSN 1331-4343. doi: 10.7153/mia-11-31.

[22] L. Lami. A solution of the generalized quantum stein's lemma. *IEEE Transactions on Information Theory*, 71(6):4454–4484, June 2025. ISSN 1557-9654. URL http://dx.doi.org/10.1109/TIT.2025.3543610.

[23] L. Lami. Generalised quantum Sanov theorem revisited. 2025. URL https://arxiv.org/abs/2510.06340.

[24] K. Landsman. *Foundations of Quantum Theory : From Classical Concepts to Operator Algebras*. Springer International Publishing : Imprint : Springer, 2017. ISBN 9783319517773.

[25] K. Landsman. Randomness? What randomness? *Foundations of Physics*, 50:61–104, 2 2020. ISSN 0015-9018. doi: 10.1007/s10701-020-00318-8.

[26] K. Landsman. Okinawa Lectures on Entropy, 2026.

[27] G. Lindblad. Completely positive maps and entropy inequalities. *Communications in Mathematical Physics*, 40(2):147–151, June 1975. doi: 10.1007/BF01609396.

[28] M. Mosonyi, Z. Szilagyi, and M. Weiner. On the error exponents of binary state discrimination with composite hypotheses. *IEEE Transactions on Information Theory*, 68(2):1032–1067, 2022. URL http://dx.doi.org/10.1109/TIT.2021.3125683.

[29] G. Murphy. *C*-Algebras and Operator Theory*. Academic Press, 2014. ISBN 9780080924960. URL https://books.google.nl/books?id=omviBQAAQBAJ.

[30] M. Müger. Topology for the working mathematician, 2022. URL https://www.math.ru.nl/~mueger/topology.pdf.

[31] J. Nötzel. Hypothesis Testing on Invariant Subspaces of the Symmetric Group, Part I - Quantum Sanov's Theorem and Arbitrarily Varying Sources. 10 2013. doi: 10.1088/1751-8113/47/23/235303. URL http://arxiv.org/abs/1310.5553http://dx.doi.org/10.1088/1751-8113/47/23/235303.

[32] T. Ogawa and H. Nagaoka. Strong Converse and Stein's Lemma in the Quantum Hypothesis Testing. Technical report, 1999.

[33] F. Rassoul-Agha and T. O. Seppalainen. *A course on large deviations with an introduction to Gibbs measures*. American Mathematical Society, 2015. ISBN 0821875787.

[34] W. Rudin. *Principles of mathematical analysis*. McGraw-Hill, 1986. ISBN 9780070542358.

[35] M. Takesaki. *Theory of Operator Algebras I*. Encyclopaedia of Mathematical Sciences. Springer Berlin Heidelberg, 2001. ISBN 9783540422488. URL https://books.google.nl/books?id=38QIwQEACAAJ.

[36] T. van Erven and P. Harremoës. Rényi divergence and kullback-leibler divergence. 4 2014. doi: 10.1109/TIT.2014.2320500. URL http://arxiv.org/abs/1206.2459http://dx.doi.org/10.1109/TIT.2014.2320500.

[37] T. Xiao, Y. Wang, Y. Xia, Z. Li, and X. Zhou. $d+1$ Measurement Bases are Sufficient for Determining d-Dimensional Quantum States: Theory and Experiment. 8 2025. URL http://arxiv.org/abs/2507.11204.

Appendix A

Selected topics in mathematics

A.1 Nets in topology

Here we provide a short section on the use of nets in topology. In general topology, nets play an identical role to sequence in first countable spaces. One can characterize the notions of closed sets, continuity of functions, and compactness in terms of nets. We begin by defining the notion of a net and what it means for a net to converge or accumulate at a point. We closely follow parts of Sections 5.1, 5.2, and 7.2 in [30].

Definition A.1.1 (Partially Ordered Set)**.** *A pair$(I, \leq)$ is a partially ordered set if $\leq$ is a binary relation satisfying*

1. *if $i_1 \in I$ then $i_1 \leq i_1$,*
2. *if $i_1, i_2, i_3 \in I$ with $i_1 \leq i_2$ and $i_2 \leq i_3$ then $i_1 \leq i_3$,*
3. *for all $i_1, i_2 \in I$ there exists an $i_3 \in I$ such that $i_1 \leq i_3$ and $i_2 \leq i_3$.*

Definition A.1.2. *If X is a topological space then a net is a function from some partially ordered set I to X written as $(x_i)_{i \in I} \subseteq X$.*

Definition A.1.3. *If $(x_i)_{i \in I}$ is a net on some topological space (X, τ) then we say the net **converges** to some point $x \in X$ if for all open $U \in \tau$ containing x, there exists an $j(U) \in I$ such that for all $i \geq j$, $x_i \in U$. We say the net **accumulates** at $(x_i)_{i \in I}$ at a point $x \in X$ if for all open $U \in \tau$ containing x and all $j \in I$ there exists an $i \geq j$ such that $x_i \in U$.*

We will find it especially useful to consider the generalization of subsequences (for sequences) to subnets (for nets). It should be noted all nets are sequences, but not all subnets of sequences are themselves subsequences. This among other non ideal properties is why subnets are sometimes avoided in topology and replaced by the concept of filters. However, for this thesis subnets are a more than sufficient tool.

Definition A.1.4 (Subnets)**.** *Let $(x_i)_{i \in I}$ and $(x_j)_{j \in J}$ be nets on the topological space X. We call $(x_j)_{j \in J}$ a subnet of $(x_i)_{i \in I}$ if there exists a map $\eta : J \to I$ with the following two properties*

1. *for all $j \in J$, $x_j = x_{\eta(j)}$,*

2. for every $i_0 \in I$ there exists a $j_0 \in J$ such that for all $j \geq j_0$, $\eta(j) \geq i_0$.

Proposition A.1.1 (Hausdorff)**.** *A topological space X is Hausdorff if and only if all converging nets $(x_i)_{i\in I} \subseteq X$ have a unique limit point $x \in X$. In Hausdorff spaces we will often write $\lim_{i\in I} x_i = x$ to indicate the net $(x_i)_{i\in I}$ converges to the unique point x.*

Proof. See [30, Proposition 5.1.22]. □

Throughout this thesis we will only be working with Hausdorff topological spaces. Below we summarize the essential characterizations of closed and compact sets and the continuity of a function in terms of nets.

Proposition A.1.2 (Accumulation points and converging subnets)**.** *A point $x \in X$ is an accumulation point of a net $(x_i)_{i\in I}$ if and only if there is a subnet $(x_j)_{j\in J}$ of $(x_i)_{i\in I}$ converging to x.*

Proof. See [30, Proposition 5.1.36]. □

Proposition A.1.3 (Closed characterization with nets)**.** *A set $V \subseteq X$ of a topological space X is closed if and only if for all nets $(x_i)_{i\in I} \subseteq V$ converging to some $x \in X$ we have $x \in V$.*

Proof. See [30, Proposition 5.1.21]. □

Proposition A.1.4 (Compact characterization with nets)**.** *A set $K \subseteq X$ of a topological space X is compact if and only if all nets $(x_i)_{i\in I} \subseteq K$ have an accumulation point $x \in K$. Or equivalently, all nets $(x_i)_{i\in I} \subseteq K$ have a subnet $(x_j)_{j\in J}$ converging to some $x \in K$.*

Proof. See [30, Proposition 7.2.2]. □

We now state a very useful result from [11].

Proposition A.1.5. *Assume X is a compact Hausdorff space. If a net $(x_i)_{i\in I}$ has a unique accumulation point x, then the net $(x_i)_{i\in I}$ converges to x.*

Proof. See [11, Proposition 2.7 Appendix A.2] □

Proposition A.1.6 (Continuity of maps characterized with nets)**.** *Let X and Y be two topological spaces. Then a function $f : X \to Y$ is continuous if and only if for every net $(x_i)_{i\in I} \subseteq X$ converging to some $x \in X$, the net $(f(x_i))_{i\in I} \subseteq Y$ converges to some $y \in Y$.*

Proof. See [30, Proposition 5.2.5]. □

A.2 Probability theory

Definition A.2.1. *Let X be a set and $\mathcal{F}$ be a collection of subsets of X. We call $\mathcal{F}$ a $\sigma-$algebra if*

1. *$X \in \mathcal{F}$*
2. *If $A \in \mathcal{F}$ then $X \backslash A \in \mathcal{F}$*
3. *If $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F}$ then, $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.*

Theorem A.2.1 (Carathéodory Extension Theorem)**.** *Let X be a non-empty set and $\mathcal{C}$ be a collection of subsets of X satisfying*

1. *$\emptyset \in \mathcal{C}$*
2. *For $A, B \in \mathcal{C}$ we have $A \cap B \in \mathcal{C}$*
3. *For all $A \in \mathcal{C}$ there exists a sequence of disjoint sets $K_1, \cdots, K_n \in \mathcal{C}$ such that $X \backslash A = \bigcup_{i=1}^{n} K_i$.*

Then we call $\mathcal{C}_0$ a semi-algebra of X. If there is a function $\mu_0 : \mathcal{C}_0 \to [0,1]$ with the property that for all disjoint $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{C}_0$ if $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{C}_0$,

$$\mu_0\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu_0(A_n), \tag{1.1}$$

then there exists a unique measure μ on the $\sigma-$algebra $\mathcal{C}$ generated by $\mathcal{C}_0$ such that μ_0 equals μ on $\mathcal{C}_0 \subseteq \mathcal{C}$.

Proof. See [2, Theorem 1.3.3]. □

Theorem A.2.2 (Radon-Nikodym)**.** *Let (X, Ω) be a measure space and $p, q \in \mathrm{Prob}(X)$. Assume $p \ll q$ there exists a Ω measurable function $\frac{\mathrm{d}p}{\mathrm{d}q} : X \to \mathbb{R}$ such that for all $A \in \Omega$*

$$p(A) = \int_A \frac{\mathrm{d}p}{\mathrm{d}q} \mathrm{d}q.$$

Proof. See [2, Theorem 4.1.1]. □

Definition A.2.2 (Lower semicontinuous map)**.** *Let $\mathcal{X}$ be a topological space, a function $I : \mathcal{X} \to \mathbb{R}$ is* ***lower semicontinuous*** *if for all $\alpha \in \mathbb{R}$ the level sets*

$$\{x \in \mathcal{X}; f(x) \leq \alpha\}, \tag{1.2}$$

are closed.

Definition A.2.3 (Large Deviation Principle)**.** *Let $(\mathcal{X}, \mathcal{G})$ be a measure space. We say a sequence of probability measures $(\mu_n)_{n\in\mathbb{N}} \subseteq \mathrm{Prob}(\mathcal{X})$ satisfies a* ***large deviation principle*** *with rate function $I : \mathcal{X} \to [0, \infty]$, if for all measurable $A \in \mathcal{G}$,*

$$-\inf_{x\in A^\circ} I(x) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(A) \leq \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(A) \leq -\inf_{x\in A^-} I(x), \tag{1.3}$$

where I is a lower semicontinuous function, and A° and A^- are the interior and closure of A.

Proposition A.2.1. *Suppose we have set $\mathcal{X}$ which is a measure space $(\mathcal{X}, \mathcal{G})$ and topological space $(\mathcal{X}, \tau)$. If $\tau \subseteq \mathcal{G}$ then a sequence of probability measure $(\mu_n)_{n\in\mathbb{N}} \subseteq \mathrm{Prob}(\mathcal{X})$ satisfies a large deviation principle with rate function I if and only if*

$$\begin{aligned} -\inf_{x\in A} I(x) &\leq \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(A), \quad \text{for all open } A \subseteq \mathcal{X}, \\ -\inf_{x\in B} I(x) &\geq \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(B), \quad \text{for all closed } B \subseteq \mathcal{X}. \end{aligned} \tag{1.4}$$

Proof. Assume $(\mu_n)_{\mathbb{N}}$ satisfies a large deviation principle with rate function I, then we trivially get both equations in (1.4). So assume the above two inequalities hold and let $C \in \mathcal{F}$ be measurable. Since $C^\circ \subseteq C \subseteq C^-$ we have

$$-\inf_{x\in C^\circ} I(x) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(C^\circ) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(C),$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(C) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(C^-) \leq -\inf_{x\in C^-} I(x).$$

□

Theorem A.2.3 (Borel-Cantelli Lemma)**.** *Let $(X, \mathcal{F}, p)$ be a measure space and $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F}$. Then*

$$\sum_{n=1}^{\infty} p(A_n) < \infty \implies p\left(\bigcup_{k\in\mathbb{N}} \bigcap_{n\geq k} A_n \right) = 0, \tag{1.5}$$

where $\bigcup_{k\in\mathbb{N}} \bigcap_{n\geq k} A_n = \{x \in X; x \in A_n \text{ infinitely often}\}$. And

$$\sum_{n=1}^{\infty} p(A_n) = \infty \implies p\left(\bigcap_{k\in\mathbb{N}} \bigcup_{n\geq k} A_n \right) = 1, \tag{1.6}$$

where $\bigcap_{k\in\mathbb{N}} \bigcup_{n\geq k} A_n = \{x \in X; x \notin A_n \text{ for at most finitely many } n\}$.

Proof. See [2, Theorem 7.2.2]. □

Proposition A.2.2 (Pinsker's Inequality)**.** *Assume A is a finite non-empty set and $d = |A|$. Let $||\cdot||$ denote the Euclidean norm, then there exists a $C > 0$ such that for all $p, q \in \mathrm{Prob}(A)$ we have*

$$||p - q|| \leq C\sqrt{D_{KL}(p||q)}. \tag{1.7}$$

Proof. See [10, Lemma 2]. □

Theorem A.2.4 (Strong Law of Large Numbers)**.** *Suppose $(Z_i)_{i\in\mathbb{N}}$ are i.i.d real valued random variables with law $p \in \mathrm{Prob}(X)$, for some measure space $(X, \mathcal{F})$. Suppose $\mathbb{E}[|Z_i|] < \infty$ and $\mathbb{E}[Z_i] = \mu \in \mathbb{R}$ for all $i \in \mathbb{N}$ then if*

$$S_n := \frac{1}{n}\sum_{i=1}^{n} Z_i, \tag{1.8}$$

we have almost surely

$$\lim_{n\to\infty} S_n = \mu. \tag{1.9}$$

Proof. See [13]. □

Theorem A.2.5 (Weak law of Large Numbers)**.** *Let $(Z_i)_{i\in\mathbb{N}}$ be an i.i.d sequence of real valued random variables, each with law p, such that $\mathbb{E}[Z_1] = \mu \in \mathbb{R}$ and $\mathbb{E}[|Z_1 - \mu|^2] < \infty$. Let $S_n = \sum_{i=1}^{n} Z_i$, then for all $\varepsilon > 0$ we have that*

$$\lim_{n\to\infty} p^n\Big(\Big|\frac{1}{n}S_n - \mu\Big| > \varepsilon\Big) = 0. \tag{1.10}$$

Proof. See [2, Theorem 8.1.1] □

Theorem A.2.6 (Central Limit Theorem)**.** *Let $(Z_i)_{i\in\mathbb{N}}$ be an i.i.d sequence of real valued random variables, each with law p, such that $\mathbb{E}[Z_1] = \mu \in \mathbb{R}$ and $\mathbb{E}[|Z_1 - \mu|^2] < \infty$. Then for all real numbers $x_1 < x_2$,*

$$\lim_{n\to\infty} p^n\big(x_1 \leq \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x_2\big) = \frac{1}{\sqrt{2\pi}}\int_{x_1}^{x_2} e^{\frac{-t^2}{2}}\mathrm{d}t. \tag{1.11}$$

Proof. See [2, Corollary 11.1.2]. □

A.3 Schur-Weyl Duality

In this section we assume the reader is familiar with basic group theory. Our goal is to give a short account of Schur-Weyl duality. Put briefly, this result provides a powerful decomposition of the vector space $(\mathbb{C}^d)^{\otimes n}$ by making use of the the irreducible decomposition of two representations; one representation of S_n, the n permutation group, and the other of $U(d)$, the space of $d \times d$ unitary complex matrices. We will state all the necessary definitions and results which we make use of in this thesis. Let $\mathrm{GL}(d)$ be the set of $d \times d$ complex invertible matrices and $U(d) \subseteq \mathrm{GL}(d)$ the set of unitary matrices.

Definition A.3.1. *Let $d \in \mathbb{N}$ and G be a group. A map $R : G \to \mathrm{GL}(d)$ which satisfies*

$$R(g_1 g_2) = R(g_1)R(g_2), \tag{1.12}$$

*is called a d-dimensional representation of G. We say R is an irreducible representation (**irrep**) if there is no non-trivial linear subspace $V \subset \mathbb{C}^d$ which is invariant under the action of R, namely if for all $|\psi\rangle \in V$, $R(g)|\psi\rangle \in V$ then $V = \{0\}$ or $V = \mathbb{C}^d$.*

For a group G consider the collection

$$\{R_\lambda : G \to \mathrm{GL}(d_\lambda); d_\lambda \in \mathbb{N}, \text{ and } R_\lambda \text{ is an irrep.}\}, \tag{1.13}$$

of all irreducible representations of G. We say two representations R_1, R_2 are equivalent if and only if there exists a invertible $M \in \mathrm{GL}(d)$ such that for all $g \in G$

$$R_1(g) = M^{-1} R_2(g) M. \tag{1.14}$$

Then with $\hat{G}$ we represent the set of equivalence classes of irreps, and with a slight abuse of notation we write $\lambda \in \hat{G}$ to represent the class containing R_λ. If G is finite then $\hat{G}$ is also finite dimensional. Now we state one of the must fundamental results in representation theory known as Maschke's Theorem.

Theorem A.3.1 (Maschke's Theorem)**.** *Let G be a finite group and $R : G \to \mathrm{GL}(d)$ a representation. Then there exists a finite collection $\{m_\lambda; \lambda \in \hat{G}\} \subseteq \mathbb{N}$ and an invertible matrix $M \in \mathrm{GL}(d)$ such that,*

$$d = \sum_{\lambda \in \hat{G}} m_\lambda d_\lambda,$$

$$M R(g) M^{-1} := \sum_{\lambda \in \hat{G}} R_\lambda(g) \otimes I_{m_\lambda}. \tag{1.15}$$

We call m_λ the multiplicity of R_λ.

This theorem can be though of as a non-commutative extension of the prime factorization theorem. We will indirectly apply Maschke's Theorem on the following to representations P and R as defined in III.5.1 and III.5.2. It is important to notice that P and R are both actions on $(\mathbb{C}^d)^{\otimes n}$, and for all $U \in U(d)$ and $\pi \in S_n$,

$$P(\pi)R(U) = R(U)P(\pi). \tag{1.16}$$

This property can be used to show a version of Schur-Weyl duality.

Theorem A.3.2. *Define*

$$\begin{aligned}\mathcal{U} &:= \text{span}_{\mathbb{C}}\{R(U); U \in U(d)\},\\ \mathcal{V} &:= \text{span}_{\mathbb{C}}\{P(\pi); \pi \in S_n\}.\end{aligned} \tag{1.17}$$

Then $\mathcal{U}' = \mathcal{V}$ *and* $\mathcal{V}' = \mathcal{U}$.

Proof. See [3]. □

We can reformulate Theorem A.3.2 into a similar but more useful result. The main idea is to explicitly label the irreducible decomposition of $P : S_n \to (\mathbb{C}^{d\times d})^{\otimes n}$ with and $R : U(d) \to (\mathbb{C}^{d\times d})^{\otimes n}$ with Young diagrams.

Proposition A.3.1 (Irreps of S_n)**.** *The classes of irreps* $\hat{S}_n$ *of* S_n *are in a one-to-one correspondence with conjugacy classes of* S_n*. If we denote* $[\pi]$ *as the set of all* n*-permutations which are conjugate to* $\pi \in S_n$*, meaning* $\pi_1 \in [\pi]$ *if there exists a* $\gamma \in S_n$ *such that* $\pi_1 = \gamma\pi\gamma^{-1}$*, then*

$$|\hat{S}_n| = \big|\{[\pi]; \pi \in S_n\}\big|. \tag{1.18}$$

It is well known that the conjugacy classes of S_n are fully described by the cyclic decomposition of permutations in S_n, which in turn are described by the partitions of the integer $n \in \mathbb{N}$. A **partition** of $N \in \mathbb{N}$ is defined as a collection of $n_1, \cdots, n_k \in \mathbb{N}$, such that

$$N = \sum_{i=1}^{k} n_i. \tag{1.19}$$

Definition A.3.2 (Young Diagram)**.** *Consider an arbitrary partition of* $n \in \mathbb{N}$ *which we denote as* $\lambda = (n_1, \cdots, n_k)$ *where* $1 \le k \le n$ *and* $n_1 \ge \cdots \ge n_k$*. We then write* $\lambda \vdash n$ *and call* λ *a Young diagram. The Young diagram* $\lambda = (n_1, \cdots, n_k)$ *is arrangement of* n *boxes into rows of different lengths. The first row contains* n_1 *boxes, the second row* n_2 *boxes, etc. With* $\lambda \vdash_d n$ *we denote the partition* $\lambda \vdash n$ *with at most* $d \in \mathbb{N}$ *parts, i.e. if* $\lambda = (n_1, \cdots, n_k)$ *then* $k \le d$*.*

The construction of a Young diagram is best understood with a simple example. Consider $n = 5$ and $\lambda = (2, 2, 1)$ then the Young diagram of λ is showcased in Figure A.1.

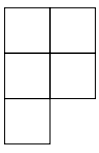


Figure A.1: Young diagram of $\lambda = (2, 2, 1)$.

It turns out that one can also label the irreps of R using Young diagrams $\lambda \vdash_d n$. This process requires a lot of extra theory, and we refer the reader to [14] for more details. To conclude we state a more suitable version of Schur-Weyl duality, which can in fact be proven using Theorem A.3.2.

Theorem A.3.3 (Schur-Weyl duality)**.** *Let $d, n \in \mathbb{N}$. Consider the two representations $P : S_n \to (\mathbb{C}^{d\times d})^{\otimes n}$ and $R : U(d) \to (\mathbb{C}^{d\times d})^{\otimes n}$. For all $\lambda \vdash_d n$, there exists an irrep R_λ of R and P_λ of P. Moreover, there exists an invertible matrix $M_{sch} \in (\mathbb{C}^{d\times d})^{\otimes n}$, called the Schur matrix such that:*

$$\begin{aligned} M_{sch}R(U)M_{sch}^{-1} &= \bigoplus_{\lambda\vdash_d n} R_\lambda(U)\otimes I_{d_\lambda} && \forall U \in U(d), \\ M_{sch}P(\pi)M_{sch}^{-1} &= \bigoplus_{\lambda\vdash_d n} I_{m_\lambda}\otimes P_\lambda(\pi) && \forall \pi \in S_n, \end{aligned} \tag{1.20}$$

where d_λ and m_λ is the multiplicity of R_λ and P_λ respectively. Equivalently,

$$(\mathbb{C}^d)^{\otimes n} \simeq \bigoplus_{\lambda\vdash_d n} \mathbb{C}^{r_\lambda}\otimes\mathbb{C}^{p_\lambda}, \tag{1.21}$$

where

$$\begin{aligned} p_\lambda &:= \dim\left(\mathrm{span}_{\mathbb{C}}\{P_\lambda(\pi)\,|v\rangle\,;\pi\in S_n, |v\rangle\in(\mathbb{C}^d)^{\otimes n}\}\right), \\ r_\lambda &:= \dim\left(\mathrm{span}_{\mathbb{C}}\{R_\lambda(U)\,|v\rangle\,;U\in U(d), |v\rangle\in(\mathbb{C}^d)^{\otimes n}\}\right). \end{aligned} \tag{1.22}$$

Proof. See [14, Theorem 5.18.1]. □

A.4 Miscellaneous results

Lemma A.4.1. *Let $n, m \in \mathbb{N}$ then*

$$\frac{m!}{n!} \geq n^{m-n}. \tag{1.23}$$

Proof. Assume that $n > m$ then,

$$n! = n(n-1)\cdots(m+1)m!,$$

but trivially $n^{n-m} \geq n(n-1)\cdots(m+1)$ and $\frac{n!}{m!} \leq n^{n-m}$. On the other hand if $n \leq m$ then

$$m! = m(m-1)\cdots(n+1)n!,$$

and $m(m-1)\cdots(n+1) \geq n^{m-n}$ so $\frac{m!}{n!} \geq n^{m-n}$. □

Lemma A.4.2. *Let $N \in \mathbb{N}$ and $\mathrm{Prob}_N(A)$ be defined as in (2.40) then,*

$$\Big(\bigcup_{N\in\mathbb{N}} \mathrm{Prob}_N(A) \Big)^{-} = \mathrm{Prob}(A), \tag{1.24}$$

where the closure is taken in the Euclidean norm topology.

Proof. Let $\varepsilon > 0$, and $p = (p_1, \cdots, p_d) \in \mathrm{Prob}(A)$ and assume without loss of generality for all $i \in \{1, \cdots, d\}$, $p_i > 0$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$ there exists an $M \in \mathbb{N}$ and $(n_1, \cdots, n_d) \in \mathbb{N}$ such that for all $i \in \{1, \cdots, d\}$,

$$\Big| p_i - \frac{n_i}{M} \Big| < \frac{\varepsilon}{2d}. \tag{1.25}$$

Define $a \in \mathbb{Q}$ as

$$a := 1 - \sum_{i=1} \frac{n_i}{M}, \tag{1.26}$$

and assume with out loss of generality that $a > 0$, then there exists an $m \in \mathbb{N}$ such that $a = \frac{m}{M}$. Next define the probability measure $q \in \mathrm{Prob}_N(A)$ as

$$q_i := \frac{n_i + m\delta_{1,i}}{M}, \tag{1.27}$$

for all $i \in \{1, \cdots, d\}$. We can confirm q is a probability measure explicitly,

$$\sum_{i=1}^{d} q_i = \frac{n_1 + m}{M} + \sum_{i=2}^{d} \frac{n_i}{M} = a + \sum_{i=1}^{d} \frac{n_i}{M} = 1, \tag{1.28}$$

and for all $i \in \{1, \cdots, d\}$, $q_i \geq 0$. By assumption

$$|a| = \Big| 1 - \sum_{i=1}^{d} \frac{n_i}{M} \Big| \leq \Big| \sum_{i=1}^{d} p_i - \sum_{i=1}^{d} \frac{n_i}{M} \Big| + \Big| 1 - \sum_{i=1}^{d} p_i \Big| \tag{1.29}$$

$$\leq \frac{\varepsilon}{2} + 0, \tag{1.30}$$

therefore

$$||p-q|| = \Big(\sum_{i=1}^{d} |p_i - q_i|^2\Big)^{\frac{1}{2}} \leq \sum_{i=1}^{d} |p_i - q_i| \leq |a| + \sum_{i=1}^{d} \Big|p_i - \frac{n_i}{N}\Big| \tag{1.31}$$

$$\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{1.32}$$

□

Proposition A.4.1. *Let $\mathcal{H}$ be a d-dimensional Hilbert space. If $I = \{1, \cdots, d+1\}$, then there exists a collection of distinct bases $(\mathcal{B}_i)_{i\in I}$ such that if $\Omega := \cup_{i\in I}\mathcal{B}_i$, then the POVM $\mu : \Omega \to B(\mathcal{H})$ defined as*

$$\mu(|x\rangle) := \frac{1}{m} |x\rangle \langle x|, \tag{1.33}$$

where $m = |\Omega|$, is an IC-POVM.

Proof. In a recent paper [37] it was shown that for a $d-$dimensional Hilbert space a collection of $d+1$ bases $(\mathcal{B}_i)_{i=0}^{d}$ is needed such that any $\rho \in \mathcal{D}(\mathcal{H})$ can be uniquely reconstructed from the set,

$$\{\langle x| \rho |x\rangle ; x \in \mathcal{B}_i, i = 0, \cdots, d\}. \tag{1.34}$$

We will present the construction of the bases and refer the reader to the paper of [37] for the rest of the proof. Fix $\mathcal{B}_0 = \{|a\rangle\}_{a=0}^{d-1}$ then consider the operator $W \in B(\mathcal{H})$ defined as

$$W := \frac{1}{\sqrt{d}} \sum_{a,b=0}^{d-1} \omega^{ab} |a\rangle \langle b|, \tag{1.35}$$

where $\omega = e^{\frac{1}{d}2\pi i}$. We claim the set $\{\Big|\psi_k^{(0)}\Big\rangle\}_{k=0}^{d-1}$ defined as

$$\Big|\psi_k^{(0)}\Big\rangle := W|k\rangle = \frac{1}{\sqrt{d}} \sum_{a=0}^{d-1} e^{ak\frac{2\pi i}{d}} |a\rangle, \qquad \forall k \in \{0, \cdots, d-1\} \tag{1.36}$$

is an orthonormal basis. Let $k, k' \in \{0, \cdots, d-1\}$ then

$$\Big\langle \psi_{k'}^{(0)} \Big| \psi_k^{(0)} \Big\rangle = \frac{1}{d} \sum_{a,b=0}^{d-1} \omega^{-ak'} \omega^{bk} \langle a| b\rangle = \frac{1}{d} \sum_{a=0}^{d-1} \omega^{a(k-k')},$$

where we use that $\overline{\omega} = \omega^{-1}$. Clearly if $k = k'$ then $\langle \psi_k^{(0)} | \psi_k^{(0)} \rangle = 1$. Now we will make use of the finite geometric series, which is for $r \in \mathbb{C}\backslash\{1\}$,

$$\sum_{i=0}^{n-1} r^i(1-r) = \sum_{i=0}^{n-1} r^i - r^{i+1} = 1 + \cdots + r^{n-1} - (r + \cdots + r^n) = 1 - r^n,$$

$$\implies \sum_{i=0}^{n-1} r^i = \frac{1-r^n}{1-r}. \tag{1.37}$$

Assume $1 \leq |k-k'| < d$ then take $r = \omega^{k-k'} \neq 1$,

$$\begin{aligned}\left\langle \psi_{k'}^{(0)} \middle| \psi_k^{(0)} \right\rangle &= \frac{1}{d} \sum_{a=0}^{d-1} \left(\omega^{k-k'}\right)^a \\ &= \frac{1}{d} \cdot \frac{1 - \left(\omega^{k-k'}\right)^d}{1 - \omega^{k-k'}} = \frac{1}{d} \cdot \frac{1 - \left(\omega^d\right)^{k-k'}}{1 - \omega^{k-k'}},\end{aligned}$$

since $(\omega^d)^{k-k'} = (e^{2\pi i})^{k-k'} = 1$ then $\langle \psi_{k'}^{(0)} | \psi_k^{(0)} \rangle = 0$ if $k \neq k'$. Define the unitary operators $U^{(r)}$ for $r = 0, \cdots, d-1$ as

$$U^{(r)} := \sum_{a=0}^{d-1} e^{i\theta_a^{(r)}} |a\rangle \langle a| , \tag{1.38}$$

where we assume for all $a, r \in \{0, \cdots, d-1\}$ that $\theta_a^{(r)} \in \mathbb{R}$ are unique. For each $i \in \{0, \cdots, d-1\}$ we can define a basis:

$$\left\{ \left| \psi_k^{(i)} \right\rangle \right\}_{k=0}^{d-1} := \left\{ R^i \left| \psi_k^{(0)} \right\rangle \right\}_{k=0}^{d-1}. \tag{1.39}$$

Now the claim is that there exists a subset $(\theta_a^{(r)})_{a,r} \subseteq \mathbb{R}$ such that the collection of bases in (1.39) insures the construct POVM μ is an IC-POVM. For the rest of the proof see [37].

□

Appendix B

Operator algebras

In this section we provide a collection of useful results from functional analysis and operators algebras. We refer the reader to [11, 29, 35, 7, 24] for a detailed treatment. Let $\mathcal{C}$ be a vector space over the field $\mathbb{C}$. We say that $\mathcal{C}$ is an **algebra** if it is equipped with an associative and distributive multiplication. Namely, for $a, b, c \in \mathcal{C}$ and $\lambda, \mu \in \mathbb{C}$ we have a product $a \cdot b = ab \in \mathcal{C}$ which satisfies

$$a(bc) = (ab)c, \quad a(b+c) = ab + ac, \quad \lambda\mu(ab) = (\lambda a)(\mu b).$$

On the algebra $\mathcal{C}$ we can define an **involution** or **adjoint** map $* : \mathcal{C} \to \mathcal{C}$ that is anti-linear and satisfies

$$(a^*)^* = a, \quad (ab)^* = b^*a^*, \quad (\lambda a + b)^* = \overline{\lambda}a^* + b^*.$$

An algebra $\mathcal{C}$ equipped with an involution is a $*$-algebra. We can further endow the underlying vector space $\mathcal{C}$ with a norm $||\cdot|| : \mathcal{C} \to \mathbb{R}^+$, turning into into a normed vector space.

Definition B.0.1 (C^* algebra)**.** *If the normed $*$-algebra $\mathcal{C}$ is a Banach space (sequentially complete) and the norm satisfies*

$$||ab|| \leq ||a|| \cdot ||b||, \qquad ||aa^*|| = ||a||^2, \tag{2.1}$$

for all $a, b \in \mathcal{C}$ then, we call $\mathcal{C}$ a C^* algebra.

Definition B.0.2 (Banach Dual)**.** *Let $\mathcal{C}$ be a C^* algebra, the **Banach dual** of $\mathcal{C}$ is defined as the set of all bounded linear functions $f : \mathcal{C} \to \mathbb{C}$. We denote the set of all linear continuous functionals with $\mathcal{C}^*$.*

Definition B.0.3 (Weak-$*$ topology)**.** *Let X be a normed vector space and for each $x \in X$ define the norm-continuous linear functional $f_x : X^* \to \mathbb{C}$. Then the **weak-$*$** topology is Hausdorff and generated by the family of semi-norm $(f_x)_{x \in X}$. Or equivalently, a net $(h_i)_{i \in I} \subseteq X^*$ converges to $h \in X^*$ if and only if for all $x \in X$*

$$\lim_{i \in I} h_i(x) = h(x). \tag{2.2}$$

We now state a fundamental result in functional analysis.

Theorem B.0.1 (Banach-Alaoglu)**.** *If Y is a normed space then the closed unit ball in Y^* is weak-$*$ compact.*

Proof. See [11, Theorem V.3.1]. □

Definition B.0.4 (Proper Ideal)**.** *Let $\mathcal{C}$ be a C^* algebra and let $I \subseteq \mathcal{C}$ be a closed linear subspace. We call I a **proper two-sided ideal** if for all $a \in \mathcal{C}$ and $b \in I$, $ab \in I$ and $ba \in I$*

Examples:

1. Let $\mathcal{H}$ be a Hilbert space. Then $(B(\mathcal{H}), *, ||\cdot||_{op})$ is a C^* algebra, where the involution is the adjoint of an operator and $||\cdot||_{op}$ is the operator norm. By definition the adjoint satisfies the involution properties so all we will prove is that the operator norm satisfies the C^* norm properties. First note the norm is sub-multiplicative

$$||ab||_{op} = \sup_{||x||=1} ||abx|| \leq \sup_{||x||=1} ||a||_{op}||bx|| = ||a||_{op}||b||_{op}. \tag{2.3}$$

   Then using Cauchy-Schwarz inequality we get

$$\begin{aligned} ||a||_{op}^2 &= \sup_{||x||=1} \langle ax, ax\rangle = \sup_{||x||=1} \langle x, aa^*x\rangle \\ &\leq \sup_{||x||=1} ||x|| \cdot ||aa^*x|| = ||aa^*||_{op} \leq ||a||_{op}||a^*||_{op}. \end{aligned} \tag{2.4}$$

   Where we can also replace a by a^* in the inequality above such that we can conclude $||a||_{op} \leq ||a^*||_{op}$ and $||a^*||_{op} \leq ||a||_{op}$, this also immediately implies that of the norm that $||aa^*||_{op} \leq ||a||_{op}^2$. So $||aa^*||_{op} = ||a||_{op}^2$ and $B(\mathcal{H})$ is a C^* algebra.

2. The $*$-sub-algebra of compact operators $K(\mathcal{H}) \subseteq B(\mathcal{H})$.

Theorem B.0.2 (Compact Operators)**.** *The $*$-sub-algebra of **compact operators** $K(\mathcal{H}) \subseteq B(\mathcal{H})$ for some Hilbert space $\mathcal{H}$ is a C^* algebra. Where an operator $a \in B(\mathcal{H})$ is called **compact** if the norm closure of the image of the unit ball*

$$B_1 := \{x \in \mathcal{H}; ||x|| \leq 1\} \subseteq \mathcal{H}, \tag{2.5}$$

is compact.

Proof. See [11, Theorem II.4.2. Proposition]. □

One particularly useful notion in operator algebras and functional analysis is a projection, which has a purely algebraic definition

Definition B.0.5. *Let $\mathcal{C}$ be a $*$-algebra, then $p \in \mathcal{C}$ is a projection if $p^2 = p = p^*$. In the case $p^2 = p$ we say p is an idempotent.*

In the case there exists a projection $\mathbf{1} \in \mathcal{C}$ such that for all $a \in \mathcal{C}$ we have $a\mathbf{1} = \mathbf{1}a$, then we call $\mathcal{C}$ unital. In unital C^* algebra's we say an operator $a \in \mathcal{C}$ is invertible if there exists an operator $a^{-1} \in \mathcal{C}$ such that $aa^{-1} = a^{-1}a = \mathbf{1}$. For unital C^* algebra's $\mathcal{C}$ we are able to define the spectrum $\sigma(a)$ of an operator $a \in \mathcal{C}$ and the set of positive operators $\mathcal{C}^+$.

Definition B.0.6 (Spectrum). *Suppose $\mathcal{C}$ is a unital C^* algebra and $a \in \mathcal{C}$. Then the spectrum of a is a subsets of $\mathbb{C}$ defined as*

$$\sigma(a) := \left\{\lambda \in \mathbb{C}; \lambda \mathbf{1} - a \text{ is not invertible.}\right\} \tag{2.6}$$

Definition B.0.7 (Positive Operator). *Suppose $\mathcal{C}$ is a unital C^* algebra and $a \in \mathcal{C}$, we say a is positive and write $a \geq 0$ if $a = a^*$ is self-adjoint and $\sigma(a) \subseteq [0, \infty)$.*

For self adjoint operators $a \in \mathcal{C}$ we have the following very useful decomposition result.

Theorem B.0.3 (Spectral Decomposition). *Let $\mathcal{C}$ be a C^* algebra and $N \in \mathcal{C}$ be a normal operator. Then there exists a spectral measure E on the Borel σ-algebra of the spectrum of N, i.e. $\sigma(N)$ such that*

$$N = \int_{\sigma(N)} z \mathrm{d}E_z. \tag{2.7}$$

Proof. See [11, Theorem IX.2.2]. □

Theorem B.0.4 (Functional Calculus). *Let $\mathcal{C}$ be a C^* algebra with identity and $a \in \mathcal{C}$. Suppose f is a holomorphic function on an open set $U \subseteq \mathbb{C}$ containing $\sigma(a) \subset U$. Then $f(a) \in \mathcal{C}$.*

Proof. See [11, Theorem VII.4.7]. □

Theorem B.0.5 (Spectral Mapping Theorem). *Let $\mathcal{C}$ be a C^* algebra with identity and $a \in \mathcal{C}$. Suppose f is a holomorphic function on an open set $U \subseteq \mathbb{C}$ containing $\sigma(a) \subset U$. Then*

$$\sigma(f(a)) = f(\sigma(a)). \tag{2.8}$$

Proof. See [11, VII.4.10]. □

In the case we take any Hilbert space $\mathcal{H}$ and consider the C^* algebra $B(\mathcal{H})$ then a projection $p \in B(\mathcal{H})$ has several nice properties.

Proposition B.0.1 (Projections in $B(\mathcal{H})$). *Let $p \in B(\mathcal{H})$ be a projection. Then we have the following:*

1. *$(\mathbf{1} - p)$ is a projection and $\ker(\mathbf{1} - p) = \mathrm{range}(p)$ so $\mathrm{range}(p)$ is necessarily closed.*
2. *$\mathcal{H} = \mathrm{range}(p) \oplus \ker(p)$.*

Proof. See [11, I 2.7. Theorem]. □

Given a net of positive real numbers $(x_\lambda)_{\lambda \in \Lambda}$ we write

$$\sum_{\lambda \in \Lambda} x_\lambda := \sup \left\{ \sum_{\lambda \in F} x_\lambda; F \subseteq \Lambda \text{ is finite} \right\}. \tag{2.9}$$

This notion can be generalized to any Banach space $(X, ||\cdot||)$. Let $(x_\lambda)_{\lambda \in \Lambda}$ be a indexed subset of X, where Λ is some arbitrary set. Denote with Λ' the set of all finite subsets of Λ. Clearly Λ' can be a partially ordered set by using inclusion as an ordering. Consider

the net $(x_F)_{F\in\Lambda'}$ defined as $x_F = \sum_{\lambda\in F} x_\lambda$ for $F \in \Lambda'$. Then if the net $(x_F)_{F\in\Lambda'}$ converges to some x_0 with respect to the norm topology on X, we write

$$\sum_{\lambda\in\Lambda} x_\lambda := x_0. \tag{2.10}$$

Proposition B.0.2. *Every Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ has an orthonormal basis E, and any vector $x \in \mathcal{H}$ can be written as*

$$x = \sum_{e\in E} \langle x, e\rangle e.$$

Moreover, there exist at most countably many $e \in E$ such that $\langle x, e\rangle \neq 0$.

Proof. See [11, Section I]. □

Now we provide a list of relevant topologies on the space $\mathcal{B}(\mathcal{H})$.

Definition B.0.8 (Topologies in $B(\mathcal{H})$)**.** *Let $\mathcal{H}$ be a Hilbert space and $B(\mathcal{H})$ the space of bounded linear operators. Then there are several topologies of interest we can apply onto $B(\mathcal{H})$:*

1. *The **weak operator topology** (wot) is defined by the family of semi-norms $(p_{x,y})_{x,y\in\mathcal{H}}$ defined as*

$$p_{x,y}(a) = |\langle a(x), y\rangle|.$$

   A net of operators $(a_\lambda)_{\lambda\in\Lambda}$ converges weakly to $a_0 \in B(\mathcal{H})$ if and only if for all $x, y \in \mathcal{H}$ we have $(p_{x,y}(a_\lambda))_{\lambda\in\Lambda}$ converges to $p_{x,y}(a_0)$ in $\mathbb{R}^+$.

2. *The **strong operator topology** (sot) is defined by the family of semi-norms $(p_x)_{x\in\mathcal{H}}$ defined as*

$$p_x(a) = ||a(x)||_{\mathcal{H}}. \tag{2.11}$$

   A net of operators $(a_\lambda)_{\lambda\in\Lambda}$ converges strongly to $a_0 \in B(\mathcal{H})$ if and only if for all $x \in \mathcal{H}$ we have $(p_x(a_\lambda))_{\lambda\in\Lambda}$ converges to $p_x(a_0)$ in $\mathbb{R}^+$.

Definition B.0.9 (von Neumann algebra)**.** *Assume $\mathbf{1} \in \mathcal{M} \subseteq B(\mathcal{H})$ is a $*$- algebra, then $\mathcal{M}$ is a von Neumann algebra if $\mathcal{M}$ is weakly closed.*

Definition B.0.10 (Commutant)**.** *For a subset $A \subseteq B(\mathcal{H})$, A' is the commutant of A and is defined as*

$$A' = \{b \in B(\mathcal{H}); ab = ba, \forall a \in A\}.$$

Now we can state an exceptionally beautiful result connecting the notions of topologies defined on operators and sets of operators satisfying purely multiplicative properties.

Theorem B.0.6 (Double Commutant Theorem)**.** *If $\mathcal{M} \subseteq B(\mathcal{H})$ is a $*$-algebra with identity then $\mathcal{M}$ is weakly closed if and only if $\mathcal{M}'' = \mathcal{M}$.*

Proof. See [29, Theorem 4.1.5]. □

We can now discuss the algebraic isomorphic between two von Neumann algebras.

Definition B.0.11 (Algebraic isomorphism)**.** *A von Neumann algebra $\mathcal{M}$ is **algebraically isomorphic** to a von Neumann algebra $\mathcal{N}$ if there is a map $\varphi : \mathcal{M} \to \mathcal{N}$ such that:*

1. *φ is a linear isomorphism between the two vector space structures on $\mathcal{M}$ and $\mathcal{N}$,*
2. *for all $m_1, m_2 \in \mathcal{M}$,*

$$\varphi(m_1 m_2) = \varphi(m_1)\varphi(m_2), \quad \varphi(m_1^*) = \varphi(m_1)^*. \tag{2.12}$$

When $\mathcal{M}$ is algebraically isomorphic to $\mathcal{N}$, we write $\mathcal{M} \simeq \mathcal{N}$. An algebraic isomorphism is both norm and σ-weakly continuous.

Let $a \in B(\mathcal{H})$. Define the operator $[a]$ to be the projection onto the norm closed linear subspace $\text{range}(a)^- \subseteq \mathcal{H}$.

Theorem B.0.7. *If $\mathcal{M}$ is a von Neumann algebra, then it contains the ranged projection of all its element. Namely if $[m]$ is the ranged projection of $m \in \mathcal{M}$ then $[m] \in \mathcal{M}$.*

Proof. See [29, Theorem 4.1.9]. □

Theorem B.0.7 indicates that projections are bountiful in von Neumann algebras, in fact the set of projections in a von Neumann algebra can completely generate the whole von Neumann algebra in two notable ways.

Theorem B.0.8. *Let $\mathcal{P}(\mathcal{M}) = \{p \in \mathcal{M}; p^2 = p = p^*\}$ be the set of all projections in a von Neumann algebra $\mathcal{M}$. Then*

$$\mathcal{M} = \text{Span}(\mathcal{P}(\mathcal{M}))^-, \tag{2.13}$$

i.e. $\mathcal{M}$ is the norm closure of the linear span of the set of all projections in $\mathcal{M}$ and,

$$\mathcal{M} = \mathcal{P}(\mathcal{M})''. \tag{2.14}$$

Proof. See [24, Theorem C.135]. □