Statistical Methods for Quantum State Tomography

Saifuddin Syed
#20387835

E-mail: saifuddin.syed@uwaterloo.ca

Abstract. In quantum state tomography we want to be able to construct an estimate $\hat{\rho}$ after performing repeated measurements on some unknown ρ. The most widely used method right now to contruct $\hat{\rho}$ is maximum likelihood estimation (MLE). MLE is flawed because it produces estimates that are implausible, and gives no bound on the errors. In this paper we introduce a relatively new approach to state tomography; Bayesian mean estimation (BME). We discuss the properties of BME and its advantages over MLE. BME is then used to construct confidence regions of $\hat{\rho}$, an analogue of confidence intervals from classical statistics.

Submitted to: *Professor Joseph Emerson for AMATH 876*

1. Introduction

Quantum tomography is the collection of methods used to characterize and validate quantum devices. In order to validate a device, we want to be able to reconstruct a state ρ by performing repeated independent measurements on N identically prepared systems. Since there are only a finite number of measurements, it is impossible to compute ρ exactly, however we can compute $\hat{\rho}$, an estimator of ρ to arbitrary precision. There are many approaches on how one should tackle this problem.

In this paper we will discuss some statistical methods to determine the best estimator for ρ. We will begin with a discussion of maximum likelihood estimators and the frequentest approach. We will see that this approach is insufficient and we must rely on Bayesian mean estimators. Once we construct our estimator, we will discuss how good the estimates are at determining the true ρ. In addition we will discuss multiple approaches to constructing these confidence regions for ρ. Open problems and next steps related to Bayesian mean estimation will be discussed.

For the purposes of this paper we will only discuss quantum state tomography as process tomography is mathematically equivalent.

2. Review of tomography

Before we begin our analysis, we need to briefly overview classical tomography as discussed in class (to standardize notation). Let $\mathcal{H} = \mathbb{C}^d$ be a Hilbert space. As discussed in class, we need $d^2 - 1$ linearly independent projective measurements in aggregate to provide sufficient information to identify a unique $\hat{\rho}$ [2]. Let $\mathcal{D} = \{D_1, \ldots, D_N\}$, be the observed data after performing N independent measurements on our unknown state ρ, where D_i is the i^{th} outcome. $\mathcal{M} = \{E_1, \ldots, E_m\}$ be all the possible outcomes. Let n_i denote the frequency of outcome E_i, and $f_i \equiv \frac{n_i}{N}$ be relative frequency. The probability of outcome E_i occurring is $p_i \equiv \mathrm{Tr}(E_i\rho)$.

For large N, $f_i \approx p_i$ due to the strong law of large numbers. So we define $\hat{\rho}_{\mathrm{tomo}}$ to be the unique solution to the following linear equations in Hilbert-Schmidt space:

$$\mathrm{Tr}(E_i\hat{\rho}_{\mathrm{tomo}}) = f_i. \tag{1}$$

This is the most natural definition of tomography because it produces a an operator that best describes the observed frequencies based on Born's rule. Although this is simple and intuitive, it is not a very good approximation. We can guarantee $\hat{\rho}_{\mathrm{tomo}}$ has unit trace, is hermitian, but we cannot ensure that the eigenvalues are non-negative (they often are). In general, as d increases, the probability of viewing a negative eigenvalue increases too. The underlying flaw with tomography is that the goal of this method is to best match relative frequencies and Born's rule, and pays no attention to positivity [4]. In addition this method does not let us quantify how bad our approximation is. Clearly we need a more sophisticated approach to tomography.

3. The Maximum Likelihood Estimator

The frequentest approach yields the intuitive and simple maximum likelihood estimator (MLE). MLE is based off idea that the best estimate of the ρ is the valid state $\hat{\rho}$ that maximizes the probability of the observed data occurring.

Definition 3.1 *We define the* ***likelihood function*** $\mathcal{L}(\rho) : \mathcal{P} \to [0, 1]$ *by*

$$\mathcal{L}(\rho) = \Pr(\mathcal{D}|\rho). \tag{2}$$

Where $\mathcal{P}$ is the set of density operators corresponding to states. The ***log-likelihood function*** *is defined to be* $\log(\mathcal{L})$*, and the* ***maximum likelihood estimator*** $\hat{\rho}_{\mathrm{MLE}}$ *is the ρ that maximizes $\mathcal{L}$ or equivalently maximizes* $\log(\mathcal{L})$*.*

By applying Born's rule and using the fact that the measurements are independent we get the following expression for $\mathcal{L}$

$$\mathcal{L}(\rho) = \Pr(\mathcal{D}|\rho) = \prod_i^N \mathrm{Tr}(D_i\rho) = \prod_j^m \mathrm{Tr}(E_i\rho)^{n_j}. \tag{3}$$

Since the likelihood function is positive, we can maximize $\mathcal{L}$ if and only if we maximize $\log(\mathcal{L}) = \sum_i^N \log(\mathrm{Tr}(D_i\rho)$. Since the $\mathrm{Tr}(D_i\rho)$ is a non-negative linear function of ρ, and the logarithm of a linear function is convex, $\log(\mathrm{Tr}(D_i\rho))$ is a convex function. Therefore the log-likelihood function is convex, since it is the sum of convex functions. This ensures that $\mathcal{L}$ has a unique local maximum which we call $\hat{\rho}_{\mathrm{MLE}}$ [4].

3.1. MLE Versus Tomography

Now that we know what $\hat{\rho}_{\mathrm{MLE}}$ is, and that it exists, we would like to determine its relationship with $\hat{\rho}_{\mathrm{tomo}}$. For MLE to be an improvement over tomography, when $\hat{\rho}_{\mathrm{tomo}}$ is a valid state, we want it to equal $\hat{\rho}_{\mathrm{MLE}}$. This leads us to the first theorem.

Theorem 3.2 [4] *If $\hat{\rho}_{tomo}$ is a valid quantum state, then $\hat{\rho}_{\mathrm{MLE}} = \hat{\rho}_{\mathrm{tomo}}$.*

Proof. Let $f_j = \frac{n_j}{N}, \hat{p}_j = \mathrm{Tr}(E_j\rho)$ as before. So we now have:

$$\log(\mathcal{L}(\rho)) = \sum_j^m n_j \log(\mathrm{Tr}(E_j\rho)) \tag{4}$$

$$= N\sum_j^m f_j \log(p_j)) \tag{5}$$

$$= N\sum_j^m [f_j \log(f_j) - (f_j \log(f_j) - f_j \log(p_j)] \tag{6}$$

$$= -N[H(f) + D(f||p)] \tag{7}$$

Where H, D are the Shannon entropy and relative entropy respectively. Since we are looking to maximize over ρ, H is irrelevant. The relative entropy is non-negative, and is uniquely zero when $f_j = p_j$, $\forall j$. This is precisely (1), the set of linear equations governing $\hat{\rho}_{\mathrm{tomo}}$. Therefore if $\hat{\rho}_{\mathrm{tomo}}$ is a valid state, then $\hat{\rho}_{\mathrm{tomo}} = \hat{\rho}_{\mathrm{MLE}}$ □

This theorem in a sense shows that MLE can be thought of as a correction to tomography. Let $\mathcal{A}$ denote the polytope in Hilbert-Schmidt space enclosed by the hyperplanes $\mathrm{Tr}(E_i\rho) = p_i = 0$. Since for $\hat{\rho}_{\mathrm{tomo}}$, satisfies, $p_i \geq 0$, we have $\hat{\rho}_{\mathrm{tomo}} \in \mathcal{A}$. Let $\tilde{\mathcal{L}}$ extend $\mathcal{L}$ to $\mathcal{A}$ in the natural way, refer to figure 1. By an identical argument as Theorem 3.2, we have $\tilde{\mathcal{L}}$ is maximized at $\hat{\rho}_{\mathrm{tomo}} = \hat{\rho}_{\mathrm{MLE}}$. If $\hat{\rho}_{\mathrm{tomo}} \notin \mathcal{P} \subset \mathcal{A}$, the maximum of $\tilde{\mathcal{L}}|_{\mathcal{P}} = \mathcal{L}$ is in $\partial\mathcal{P}$ (since $\mathcal{P}$ is compact). Therefore if $\hat{\rho}_{\mathrm{tomo}}$ is not a valid quantum state we have $\hat{\rho}_{\mathrm{MLE}}$ is on the boundary of $\mathcal{P}$. Thus $\hat{\rho}_{\mathrm{MLE}}$ will be rank deficient, and will have some eigenvalues that are zero. [4]

In a sense MLE can be considered tomography restricted to the valid quantum states. Hence MLE is truly an improvement over tomography in terms of estimation. However, the fact that MLE gives rise to eigenvalues that are zero is very troublesome.

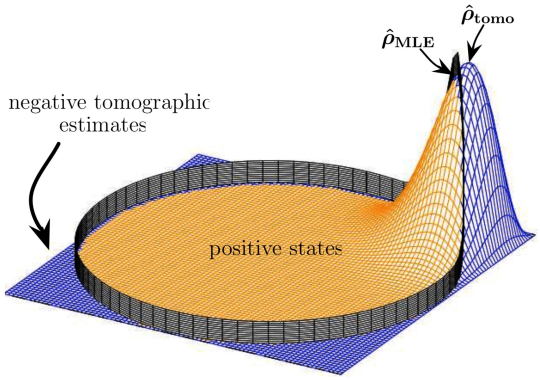

Figure 1. [4] In our case $\mathcal{P}$ is the yellow circle, and $\mathcal{A}$ would be the entire surface. $\mathcal{L}$ and $\tilde{\mathcal{L}}$ coincide on $\mathcal{P}$. This image shows the case where $\hat{\rho}_{\text{tomo}} \notin \mathcal{P}$, thus $\hat{\rho}_{\text{MLE}}$ is on the boundary of $\mathcal{P}$

3.2. Zero Eigenvalues and Their Flaws

The problem with tomography and MLE is that both of those approaches are fundamentally flawed. Both of those approaches interpret frequencies as probabilities, and construct a state that best fits those probabilities. However the purpose of a quantum state is to make predictions for future experiments, and not just explain the observed data. The estimator we find should encode the knowledge we have about the system, and the knowledge of the past is not necessarily the best description for the future.

A rather extreme but relevant example would be coin flipping. Suppose you flipped a coin N times and each time you flipped you got heads. This does not imply the probability viewing a tails is zero, but if we viewed the MLE to compute the probability of a tails occurring it would be zero. This shows that we need a more sophisticated approach that will not only give us a state that well approximates the observed data but one that also gives us a plausible one.

Similarly, since eigenvalues correspond to probabilities, having an eigenvalue that is zero implies that you are absolutely certain that a particular state will never occur, which is far too strong a conclusion to make after a finite number of experiments. [4]

An other problem is that zero eigenvalues are incompatible with any error bars. Since probabilities are are always greater than or equal to zero, any error bar on the eigenvalue will have a negative range.

4. Bayesian Mean Estimators

Bayesian methods have become popular in the last few decades. The Bayesian philosophy is that the best estimate for ρ is the average of all states consistent with the observed data weighted by their likelihood.

The procedure to determine the Bayesian mean estimator (BME) is as follows:

(i) Use the given data to construct the likelihood function $\mathcal{L}$ as defined previously

(ii) Choose a probability distribution over $\mathcal{P}$, $\pi_0(\rho)d\rho$ which we will call the **prior distribution**

(iii) Define a new probability distribution over $\mathcal{P}$ called the **posterior distribution** given by the formula

$$\pi_f(\rho)d\rho = C\mathcal{L}(\rho)\pi_0(\rho)d\rho, \tag{8}$$

where C is a normalizing factor

(iv) We define the $\hat{\rho}_{\text{BME}}$ by the expectation value of ρ with respect to the posterior distribution

$$\hat{\rho}_{\text{BME}} \equiv \langle\rho\rangle_{\pi_f} = \int_{\mathcal{P}} \rho\pi_f(\rho)d\rho. \tag{9}$$

The Bayesian mean estimator takes into account not only the optimal state, but rather all the states that could produce the observed data. This gives a more conservative and reliable estimate.

For example suppose that after N flips of a coin, n heads occur with probability p. In this case the likelihood function is

$$\mathcal{L}(p) = p^n(1-p)^{N-n}, \tag{10}$$

and the MLE in this case is

$$\hat{p}_{MLE} = \frac{n}{N}. \tag{11}$$

If we observe no heads(tails) then the MLE says that the probability of heads(tails) occurring is zero; which is an unrealistic statement.

With the a Bayesian approach, we need to chose a prior. If we choose uniform distribution with respect to the Lebesgue measure i.e. $\pi_0(p)dp = dp$, then we get

$$\hat{p}_{BME} = \frac{n+1}{N+2}. \tag{12}$$

Since n is non-negative we get the the BME is always positive. When we have no information (i.e. $N = n = 0$), we have that the probability of either heads or tails occurring is one half, which one should expect given no information. Also if we get all heads(tails), the estimator does not rule out the possibility of getting tails(heads), it just makes their probability of occurring to be less. This property that BME has is very useful for determining a plausible quantum state. This is because the probabilities of unobserved events occurring are non-zero. [4]

4.1. Robustness of Prior Distributions

An issue we have to deal with is the choice of the prior. If we had chosen the prior to be $d\pi_0(p) = \frac{1}{2}(\delta(p) + \delta(1-p))$, then if one tail were to occur then $\hat{p}_{BME}$ would be zero, which we do not want. Furthermore if both heads and tails are observed then the posterior vanishes completely and is no longer a valid probability. The problem is that the prior assigns zero probability to a possible set of observable events.

Definition 4.1 *We define a prior to be* ***fragile*** *if the following 3 equivalent conditions occur:*

(i) π_0 *assigns zero probability to a finite-length measurement.*

(ii) $\exists$ *measurement record* $\mathcal{D}$ *such that the Bayesian estimation using* π_0 *will result in a zero probability.*

(iii) $\exists$ *a measurement record* $\mathcal{D}$ *that will annihilate* π_0 *and the Bayesian estimation fails completely.*

We define a prior to be ***robust*** *if it is not fragile. So by definition robust a prior guarantees a full rank estimate.*

So continuing the above example we see that $\pi_0(p)dp = dp$ is robust, and $d\pi_0(p) = \frac{1}{2}(\delta(p) + \delta(1-p))$ is fragile.

An example of a robust prior in the quantum setting, is the Hilbert-Schmidt measure, which is a probability measure on $\mathcal{P}$, with the distribution being the uniform Lebesgue measure on the Hilbert-Schmidt space. Each observation rules out at most one state. For example if a $|0\rangle\langle 0|$ observed then $|1\rangle\langle 1|$ can not be the true state. Since there are an uncountable number of states, there is no finite length sequence of measurements that could rule out every possible state. Thus the Hilbert-Schmidt prior is robust.

The following theorem, provides a sufficient condition for robustness

Theorem 4.2 [4] *A prior with support on a smooth curve in atleast* $(d-1)^2$ *dimensions is robust.*

In general most priors that a good experimentalist would intuitively chose are robust. From now on, unless stated otherwise we will assume that the priors we work with are robust.

4.2. Natural error bars

One of the major flaws with MLE was that it produced zero eigenvalues, hence we could not always put a bound on the error in the eigenvalues. Although BME ensures the eigenvalues of the estimator are positive, that is not enough to say there are meaningful estimate for the error. First we need to define what we mean by error. Intuitively we want $\Delta\rho$ such that

$$\rho = \hat{\rho} \pm \Delta\rho, \tag{13}$$

by which we mean $\rho_{ij} = \hat{\rho}_{ij} \pm \Delta\rho_{ij}$. However, (13) is a poor way to approach this problem as this assumes the entries of the matrix are independent of each other. We will discuss in further detail in section 5 how to construct the Bayesian equivalent of confidence intervals, confidence regions. We can however put a bound on the error of the eigenvalues of $\hat{\rho}_{\mathrm{BME}}$.

Given an observale X and $\rho \in \mathcal{P}$, we define the expectation value of X with respect to ρ to be

$$\langle X \rangle_\rho \equiv \mathrm{Tr}(X\rho), \tag{14}$$

Then the expected error of X is the variance of the expectation value of X, ie $\Delta \langle X \rangle^2$. We can write this as

$$\Delta \langle X \rangle^2 = \int \langle X \rangle_\rho^2 \pi_f(\rho) d\rho - \left[\int \langle X \rangle_\rho \pi_f(\rho) d\rho \right]^2. \tag{15}$$

Since $\langle X \rangle$ parametrizes one dimension of the Hilbert-Schmidt space, to compute the distribution of $\langle X \rangle$ we can we can integrate over the the other $d^2 - 2$ dimensions denoted by σ. Since $d\rho = d\sigma d\langle X \rangle$ we have $\pi_f(\langle X \rangle) d(\langle X \rangle) = \int_\sigma \pi_f(\rho) d\rho$ by Fubini's theorem. This gives us:

$$\Delta\langle X \rangle^2 = \int \langle X \rangle^2 \pi_f(\langle X \rangle) d\langle X \rangle - \left[\int \langle X \rangle \pi_f(\langle X \rangle) d\langle X \rangle \right]^2. \tag{16}$$

If $|\lambda\rangle$ is an eigenvector of $\hat{\rho}_{BME}$ with eigenvalue λ then if we let $X = |\lambda\rangle\langle\lambda|$, we get $\langle X \rangle = \lambda$. Hence $\Delta\lambda^2 = \Delta\langle X \rangle^2$, since $0 \leq \lambda \leq 1$, and $\pi_f(\lambda) d\lambda$ is a probability distribution on the interval $[0, 1]$. By classical probability theory [3]

$$\Delta\lambda^2 \leq (\lambda_{\max} - \mathbb{E}(\lambda))(\mathbb{E}(\lambda) - \lambda_{\min}) = (1 - \lambda)\lambda. \tag{17}$$

This bound is saturated by the $\pi_0(\lambda) = (1 - p)\delta(\lambda) + p\delta(1 - \lambda)$.

In practice, well behaved priors produce convex posteriors, in which case $\Delta\lambda \leq \lambda$. Thus every eigenvalue bounds its own uncertainty. Thus BME also gives reasonable bounds on it's eigenvalues. [4]

4.3. Optimality of BME with respect to operational divergence

It is not enough that Bayesian mean estimators give plausible results, for them to be applicable they need to produce accurate results as well. We want $\hat{\rho}_{\mathrm{BME}}$ to be as "close" to ρ as possible. We also want this measure of "closeness" to be physical. This brings us to the following definition.

Definition 4.3 *An **operational divergence** $\Delta(\rho : \hat{\rho})$, is a measure of how well $\hat{\rho}$ estimates ρ where:*

(i) Δ represents an outcome of a physically implemented process.

(ii) If $\hat{\rho}_1$ is a better estimate than $\hat{\rho}_2$, then $\Delta(\rho : \hat{\rho}_1) < \Delta(\rho : \hat{\rho}_2)$.

(iii) $\hat{\rho} = \rho$ is the best estimate of ρ, i.e. $\Delta(\rho : \rho) < \Delta(\rho : \hat{\rho})$, $\forall \hat{\rho}$.

It turns out that BME is not only an accurate estimate ρ, it is the *most* accurate estimate of ρ. This is summarized in the following theorem.

Theorem 4.4 *Let Δ be an operational divergence for ρ. For any N, if we have procedure that produces an estimate $\hat{\rho}$ for ρ after N measurements, then*

$$\Delta(\rho : \hat{\rho}) > \Delta(\rho : \hat{\rho}_{\text{BME}}). \tag{18}$$

What makes this a remarkable result is that it is true for any N, not just asymptotically. This result does not mean that $\hat{\rho}_{\text{BME}}$ will be accurate, but it does mean that it will be the most accurate result you will get. Another point of note is that BME optimizes accuracy in terms of operational divergence, but it may not be the most accurate for other measures of estimation such as trace-distance or fidelity. [4] [6]

5. Confidence Regions

In classical statistics, once you have estimated a quantity, you want to know how reliable your estimate is. For a point estimate, one of these measures is confidence intervals. A confidence interval allows us to determine possible intervals which contain the true value of the quantity is question, with some fixed probability. This notion can be extended to the quantum setting. It is natural to ask what regions of $\mathcal{P}$ contain ρ with some probability α. Bayesian mean estimators allow us to construct such regions, and we can construct them so that they are independent of prior.

We begin our analysis of such regions by the following definition:

Definition 5.1 *Let $0 \leq \alpha \leq 1$, $A \subset \mathcal{P}$. We define the **coverage probability** of A to be,*

$$Pr(\rho \in A). \tag{19}$$

*If $Pr(\rho \in A) \geq \alpha$ we say that A is a **confidence region** of ρ with confidence α.*

Finding a confidence region is not hard, for example $\mathcal{P}$ is always a confidence region that contains ρ, hence $\mathcal{P}$ is a confidence region of ρ is confidence 1. Clearly, this is not very useful information since does not tell us anything new about the ρ. In general there are uncountably many confidence regions, the difficulty is find the one that gives us the most information about where ρ is located in the Hilbert-Schmidt space. This intuitively would be the smallest such confidence region with confidence α.

Another issue we want to avoid is the choice of prior. For a confidence region, coverage probability of should not depend on prior since there is often debate about which prior is the best, and different priors can produce drastically different results.

Given observed data $\mathcal{D}$ we want to construct a confidence region $\hat{\mathcal{R}}(\mathcal{D})$ such that the coverage probability of $\hat{\mathcal{R}}(\mathcal{D})$ is at least some specified confidence level α.

Remark 5.2 First point worth mentioning is that the following are not equivalent [5]:

(i) $Pr(\rho \in \hat{\mathcal{R}}(\mathcal{D})) \geq \alpha$,

(ii) $Pr(\rho \in \hat{\mathcal{R}}(\mathcal{D})|\mathcal{D}) \geq \alpha$.

(i) says "the probability that $\hat{\mathcal{R}}(\mathcal{D})$ will contain ρ is atleast α", whereas (ii) says "given the observed data, the probability that $\hat{\mathcal{R}}(\mathcal{D})$ will contain ρ is atleast α." This is a subtle point, but (i) implies that $\hat{\mathcal{R}}$ is predetermined in the sense that once $\mathcal{D}$ is observed $\hat{\mathcal{R}}(\mathcal{D})$ is fixed, thus is independent of prior. However (ii) requires you to choose a prior in order to compute. (ii) corresponds to credibility regions.

Definition 5.3 *Let $0 \leq \alpha \leq 1$, $A \subset \mathcal{P}$, and let $\mathcal{D}$ be observed data. Then we say A is a **credibility region** with credibility α if*

$$Pr(\rho \in A|\mathcal{D}) \geq \alpha. \tag{20}$$

By the above remark, to get a credible region, we need to choose a prior [1].

It should also be noted that confidence regions do not provide probabilistic data about any particular experiment. So if we have our $\hat{\rho}_{\text{BME}}$, it is important to realize that α is not the probability of success, it is a measure of how confident we are in our estimator as a whole. [5]

In this paper we will discuss two methods to construct such regions. Both were independently proposed within a few months of each other. Robin Blume-Kohout's approach was to use likelihood ratio estimates, whereas Christandl and Renner used Bayesian credible regions for a Hilbert-Schmidt prior and enlarged said regions in a specific way.

5.1. Likelihood Ratio Regions

Definition 5.4 *Given observed data $\mathcal{D}$, we define the **log likelihood ratio** $\lambda : \mathcal{P} \to [0, \infty)$ by*

$$\lambda(\rho) = -2 \log \left(\frac{\mathcal{L}(\rho)}{\mathcal{L}(\hat{\rho}_{\text{MLE}})} \right). \tag{21}$$

$\lambda(\rho)$ is a measure of how close ρ is to $\hat{\rho}_{\text{MLE}}$, where $\rho = \hat{\rho}_{\text{MLE}}$ if and only if $\lambda(\rho) = 0$. The negative factor is is there to make λ positive and the 2 is there due to convention from classical probability theory.

Definition 5.5 *Given observed data $\mathcal{D}$, and $0 \leq \alpha \leq 1$ we define $\hat{\mathcal{R}}_\alpha(\mathcal{D})$ to be the **likelihood ratio region** with confidence α, where*

$$\hat{\mathcal{R}}_\alpha(\mathcal{D}) = \{\rho|\lambda(\rho) < \lambda_\alpha(\rho)\}. \tag{22}$$

*The **threshold**, $\lambda_\alpha(\rho)$ depends on α and the dimension of the Hilbert space, d.*

Clearly the threshold $\lambda_\alpha(\rho)$ plays an integral part in the construction of the confidence regions. As λ_α gets larger $\hat{\mathcal{R}}_\alpha(\mathcal{D})$ gets bigger, so we want to chose $\lambda_\alpha(\rho)$ such that it is the smallest threshold that makes $\hat{\mathcal{R}}_\alpha(\mathcal{D})$ a confidence region with coverage probability α.

The main difficulty is determining what $\lambda_\alpha(\rho)$ is. The precise value of $\lambda_\alpha(\rho)$ depends on the confidence level α and dimension of the Hilbert space, and there is currently no closed form available. We can however approximate it.

Blume-Kohout shows that $\lambda_\alpha(\rho)$ is "almost" independent of ρ, in the sense that the values fluctuate slightly as ρ changes. In order to remove the ρ dependence, we want to find a λ_α such that

$$\lambda_\alpha \geq \lambda_\alpha(\rho) \quad \forall \rho \in \mathcal{P}. \tag{23}$$

Clearly λ_α will not be ideal, since coverage probability and region size increase as λ_α increases, so we want to find the smallest value of λ_α to ensure we gain as little region size as possible. By definition of $\hat{\mathcal{R}}_\alpha$ we have $\rho \in \hat{\mathcal{R}}_\alpha$ iff $\lambda(\rho) < \lambda$.

Definition 5.6 *We define the* ***complementary cumulative distribution function*** *by*

$$F(\lambda_\alpha|\rho) = Pr(\lambda(\rho) > \lambda_\alpha|\rho). \tag{24}$$

To get a valid λ we want to solve $\sup_\rho F(\lambda_\alpha|\rho) = 1 - \alpha$. Again finding F explicitly is a difficult task, however we can find upper bounds.

Theorem 5.7 *For any data $\mathcal{D}$, for any measurement of N copies of d-dimensional systems we have*

$$F(\lambda_\alpha) \leq N^{d^2-1} e^{-\lambda_\alpha/2}. \tag{25}$$

In the case where $\mathcal{D}$ is obtained from independent measurements of identically prepared systems, one has

$$F(\lambda) \leq \frac{\gamma(k/2, \lambda_\alpha/2)}{\Gamma(k/2)} + e^{-\lambda_\alpha/2}\left[\left(1 + \frac{\sqrt{3e\lambda_\alpha}}{\pi}\right)^k - \frac{\sqrt{3e\lambda_\alpha}}{\pi}^k\right]. \tag{26}$$

Where k is the number of linearly independent observables measured, γ is the upper incomplete gamma function, and Γ is the gamma function.

(26) is slightly stronger than (25), but also assumes a stronger result. One can solve for λ_α numerically to obtain confidence regions. The proof is quite long, hence it is omitted. Since we are estimating an upper bound for $\lambda_\alpha(\rho)$ we will get slightly larger confidence regions than needed. [5]

5.2. Bayesian Credible Regions from Hilbert-Schmidt Prior

Christandl and Renner's approach is very different than the likelihood approach used by Blume-Kohout which I will outline now. They start off with a slightly more general set up. Let $\mathcal{S}_1, \ldots, \mathcal{S}_N$ be identical finite dimensional quantum systems with associated Hilbert spaces $\mathcal{H} = \mathbb{C}^d$. This collection is measured according to an arbitrary POVM $\{B_n\}_n$ on $\mathcal{H}^{\otimes n}$, where $\sum_n B_n = \mathbf{1}_{\mathcal{H}^{\otimes n}}$. The POVM elements correspond to outcomes resulting from the measurement on the n systems. This is more general because we are not assuming that the measurements are independent.

For each B_n we can define a probability density

$$\mu_{B_n}(\rho)d\rho = C_{B_n} \operatorname{Tr}[B^n \rho^{\otimes n}]d\rho. \tag{27}$$

Where $d\rho$ is the Hilbert-Schmidt measure, as defined previously. $\mathcal{L}(\rho) = \operatorname{Tr}[B^n \rho^{\otimes n}]$ is the likelihood function as defined previously. Analogous to BME, μ_{B_n} can be considered the posteriori distribution with the Hilbert-Schmidt prior. Since this approach does not use any properties of MLE or BME we will just refer to it as μ_{B_n}. This is sufficient information to compute the confidence regions.

Theorem 5.8 (Confidence Regions From Credible Regions) [7] *Let $\alpha \in [0, 1]$. For all B_n, let $\Gamma_{\mu_{B_n}} \subset \mathcal{P}$ be a set of states satisfying*

$$\int_{\Gamma_{\mu_{B_n}}} \mu_{B_n}(\rho)d\rho \geq 1 - \frac{(1-\alpha)}{2\binom{2n+d^2-1}{d^2-1}} \geq \alpha \tag{28}$$

then we get the following confidence region $\Gamma^{\delta}_{\mu_{B_n}}$ with confidence α, as in

$$\Pr_{B_n}(\rho \in \Gamma^{\delta}_{\mu_{B_n}}) \geq \alpha. \tag{29}$$

$\Pr_{B_n}$ refers to the distribution of measurement outcomes B_n when measuring $\rho^{\otimes n}$ (i.e., outcome B_n has probability $\operatorname{tr}[B_n \rho^{\otimes n}]$). Finally,

$$\Gamma^{\delta}_{\mu_{B_n}} = \{\rho | \exists \rho' \in \Gamma_{\mu_{B_n}} \textit{ such that } F(\rho, \rho')^2 \geq 1 - \delta^2\} \tag{30}$$

where $\delta^2 = \frac{2}{n}\left[\ln\left(\frac{2}{1-\alpha}\right) + 2\ln\binom{2n+d^2-1}{d^2-1}\right]$ and $F(\rho, \rho') = \|\sqrt{\rho}\sqrt{\rho'}\|_1$ is the fidelity.

Thus is we can find any region satisfying (28), then we automatically get a confidence region. Again the proof is omitted due to its length. What this theorem says is that given a credible region $\Gamma_{\mu_{B_n}}$ for the Hilbert-Schmidt prior satisfying (28), we can enlarge it to a confidence region by (29).

5.3. Differences in Approach

Both approaches have their advantages and drawbacks. Although Christandl and Renner's approach to constructing confidence regions is very different from MLE/BME, their technique is not unrelated. As stated previously, μ_{B_n} is the posteriori distribution

with the Hilbert-Schmidt prior. This implies that their method is nearly optimal due to the efficiency of BME.

We also have that their approach is in a sense more general than Blume-Kohout's because they do not assume independent measurements, which experimentally may be too strong of an assumption. Finally it has the added advantage of producing a credibility region as well as a confidence region for the Hilbert-Schmidt prior as opposed to just a confidence region in the likelihood ratio case.

Their method is not perfect because there are a lot of questions left to be answered. First of all, their method tells us if we can find a region satisfying (20) then we have a confidence region, however it is unclear of how to find such a subset, let alone how to decide the optimal one. The obvious choice would be to pick the smallest of all the $\Gamma_{\mu_{B_n}}$ satisfying (28), however it is not an easy task in general to determine what those are. A related flaw is verifying (28) can be a difficult task in general since computing integrals over P in Hilbert-Schmidt space is not easy; we will discuss this issue in more detail in the next section.

Finally Christandl, and Renner's method offers no obvious way to compare the strength this procedure has compared to other methods. By the sheer fact that there method is related to BME's indicates that is is efficient, but it is unclear how much better it is than other methods, if at all.

In contrast Blume-Kohout's method with likelihood region estimators is not as general since he assumes independent measurements, however his method is far more intuitive in approach. The likelihood ratio is a measure of how far a state deviates from $\hat{\rho}_{\mathrm{MLE}}$, and the likelihood region we defined says states that are close to $\hat{\rho}_{\mathrm{MLE}}$ are more likely to be the true state, than those far away.

Some advantages of likelihood ratio regions include the following:

- They are optimized to give the smallest regions of any region estimator, and also offer the most nearly-optimal worst-case behaviour. By smallest region, we mean smallest volume with respect to *any* measure $d\rho$ on $\mathcal{P}$ [5].
- They are convex, because λ is convex, thus they can be manipulated using convex programming [5].
- It is easy to determine whether or not a state ρ is in the the confidence region; one simply computes $\lambda(\rho)$ and compares an inequality.

The main issue is the likelihood regions are incredible difficult to compute precisely and need to be estimated. Depending on how good the estimate of λ_α is, one can lose quite a bit of efficiency. Another issue is that although we can easily determine if a state is contained in the likelihood region, we do not have an explicit formula for $\mathcal{R}_\alpha(\mathcal{D})$ so performing manipulations can be difficult.

Overall the Bayesian credible region method proposed by Christanl and Renner is more general, however the likelihood ratio method proposed by Blume-Kohout is easy to implement. Even though Christanl and Renner's philosophy is quite different than

Blume-Kohout, their techniques are very similar, in the sense that both use Bayesian methods in some way.

6. Criticisms of BME

Although Bayesian mean estimators are versatile, intuitive and offer the most accurate estimate of ρ, they are not perfect. There are still some fundamental challenges that need to be worked out.

6.1. Finding a Suitable Prior is Hard

One of the most important questions related to BME is "what is the best choice of prior,"and "what is the penalty for choosing a wrong prior?". The optimality of $\hat{\rho}_{\text{BME}}$ depends on how close the prior is to the true distribution of the unknown states. It is unclear how bad the estimate can possibly be if the prior is not optimized for the particular scenario. Even worse, it is unclear if choosing the prior that best matches the true distribution is even robust or not.

In experiments, one wants to try and remain as impartial as possible, hence we want the prior that shows the least amount of bias. Finding the most uninformative prior is still an open question, since there is no natural choice. All of these are very important questions that need to be discussed. [4]

6.2. Integration is Hard

MLE boils down to optimizing the likelihood function $\mathcal{L}$. Computationally speaking this is not that difficult of a task, since optimization has been studied quite extensively. In contrast constructing $\hat{\rho}_{\text{BME}}$ relies almost purely on integration. One advantage MLE has over BME is that there are far more efficient algorithms available for MLE than BME, since MLE is established. Algorithms for BME began developing fairly recently so there is likely to be much faster methods as time goes on.

Another issue with BME is that even if one knows how to integrate, actually defining the boundary can be difficult. Unlike classical probability where we tend to integrate over simplices, in the quantum setting we often want to integrate over curved regions of Hilbert-Schmidt space. Defining the boundary analytically is challenging.

It should be noted that although there is much difficulty, actually computing $\hat{\rho}_{\text{BME}}$ can be done. Blume-Kohout proposed an implementable quantum version of the Metropolis-Hastings algorithm to compute $\hat{\rho}_{\text{BME}}$. The Metropolis-Hastings algorithm is a commonly used algorithm in classical Bayesian estimation. It can be thought of as a biased Monte Carlo simulation. [4]

6.3. Working With Large Systems is Hard

The final concern about Bayesian mean estimation is that it does not play nice with large systems. Quantum devices can only provide coherent control over 8 to 12 quibit

systems. Within the next 5 years that range is expected to jump to twenty to thirty. As the dimension of a Hilbert space increases, performing state estimation becomes more infeasible. For example the Hilbert space of a 30 quibit quantum register requires just under 1 million terabytes of data to store one matrix. As time goes on we need to build better hardware, and implement better techniques. The ideas presented in this paper are still immature and much more work needs to be done. [4]

7. Conclusion

As quantum devices become more sophisticated, the need for better estimation tools increases as well. We saw that the model used in standard tomography is far too simplistic to give a good estimate of ρ. This led to the define the likelihood function and the maximum likelihood estimator. Although $\hat{\rho}_{\mathrm{MLE}}$ was a better estimate than $\hat{\rho}_{\mathrm{tomo}}$, it was inherently flawed because it produced eigenvalues that are zero, and did not give a natural bound on how bad the estimate was. We then approached the problem of quantum state tomography using Bayesian mean estimators. We showed that BME provided plausible and accurate states, with natural bounds on its error. Finally we discussed how one can use the techniques related to Bayesian mean estimators and construct confidence regions ρ independent of a prior distribution. Bayesian mean estimation is very new but in time it can be a very powerful tool for experimentalists.

[1] *The Bayesian Choice - A Decision Theoretic Motivation*. Springer-Verlag, 1994.
[2] *Theory of Open Quantum Systems with Applications*. University of Waterloo, 2013.
[3] Rajendra Bhatia and Chandler Davis. A better bound on the variance. *The American Mathematical Monthly*, 107(4):pp. 353–357, 2000.
[4] Robin Blume-Kohout. Optimal, reliable estimation of quantum states. *New Journal of Physics*, 12(4):043034, 2010.
[5] Robin Blume-Kohout. Robust error bars for quantum tomography. *arXiv preprint arXiv:1202.5270*, 2012.
[6] Robin Blume-Kohout and Patrick Hayden. Accurate quantum state estimation via" keeping the experimentalist honest". *arXiv preprint quant-ph/0603116*, 2006.
[7] Matthias Christandl and Renato Renner. Reliable quantum state tomography. *Phys. Rev. Lett.*, 109:120403, Sep 2012.