Maximum a posteriori estimation of quantum states

Vikesh Siddhu
Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, U.S.A.
Date: 30th May’18

Abstract

Using a Bayesian methodology, we introduce the maximum a posteriori (MAP) estimator for quantum state and process tomography. The maximum likelihood, hedged maximum likelihood, maximum likelihood-maximum entropy estimator, and estimators of this general type, can be viewed as special cases of the MAP estimator. The MAP, like the Bayes mean estimator includes prior knowledge, however for cases of interest to tomography can take advantage of convex optimization tools making it numerically easier to compute. We show how the MAP and other Bayesian quantum state estimators can be corrected for noise produced if the quantum state passes through a noisy quantum channel prior to measurement.

Contents

1 Introduction 1
2 Quantum State Tomography 2
3 Bayesian Estimators 3
   3.1 Bayes Rule 3
   3.2 MAP estimate 3
   3.3 MAP and Quantum Process Tomography 4
4 Modelling Noise 5
5 Conclusion 5

1 Introduction

The task of quantum state tomography is to estimate a quantum state or density operator by performing measurements. Its classical analogue is to estimate the parameters of a probability distribution by sampling from it several times. Quantum process tomography deals with the estimation of a noisy quantum channel or completely positive trace preserving map; its classical analogue is the estimation of a conditional probability distribution.

An estimator is a procedure which uses the data from measurements to construct an estimate of the object of interest, that object is called the estimand. A point estimator provides a single best guess of the estimand; for example guessing the bias of a coin by flipping it several times, or locating a point in the qubit Bloch ball by measuring many identically prepared qubits. An interval estimator, more generally a set estimator, provides a set of plausible values for the estimand; for example a confidence interval for the bias of a coin, or a confidence region in the Bloch ball for a qubit state. A lot of effort has been devoted to constructing good estimators for quantum states. Various point estimators [1–4] and interval estimators [5–7] have been proposed.

In this work using Bayesian methods we introduce the maximum a posteriori (MAP) point estimator. One begins with a prior probability density on the set of quantum states, and using the experimental data the prior is updated to obtain a posterior density. The maximum of the posterior density gives the MAP estimate. The mean of the posterior density gives what is called the Bayes mean estimator (BME) [8]. We show that in many cases of interest to quantum tomography, the MAP estimator can be computed efficiently using convex optimization tools. This may be computationally simpler than computing the BME, which is evaluated by numerically integrating over the set of density operators.

Several well known estimators, in particular the maximum likelihood estimator (MLE) [1], the hedged maximum likelihood estimator (HMLE) [2], and the maximum likelihood-maximum entropy (MLME) estimator [3], can be viewed as special cases of the MAP estimator corresponding to particular choices of the
prior probability density. Thus the MAP estimator provides a systematic framework for discussing estimators of this general type, and casts them in a new light. In addition we show how MAP estimators can be applied to quantum process tomography.

Experimental setups are noisy, and it is useful to be able to correct the experimental data for noise. The MAP approach provides an easy way to do this if the noise can be represented by a quantum channel with known parameters (that may have been determined in a separate experiment), through which the system of interest passes on its way to the measurement device.

The rest of this paper is organized as follows. Section 2 is devoted to a general discussion of quantum state tomography and the linear inversion estimator; the material here is not new, but helps understand the later material. Section 3 introduces Bayesian estimators and the MAP estimator for quantum state and process tomography. In Sec. 4 we discuss how the simple noise model mentioned above can be incorporated in the MAP estimation framework. Section 5 contains a brief summary.

2 Quantum State Tomography

Let $H$ be a $d$-dimensional Hilbert space, and $L(H)$ be the space of operators on $H$. The set $S$ of quantum state i.e. density operators forms a convex subset of $L(H)$. Measurements on quantum systems can be described using a POVM (positive operator-valued measure), a collection of positive operators that sum to the identity in $H$. Let $\{\Lambda_i\}$ be a POVM, and $\rho$ be a density operator. The probability $p_i$ of observing an outcome corresponding to the operator $\Lambda_i$ is

$$p_i = \text{Tr}(\rho \Lambda_i).$$

One simple measurement scheme for doing quantum state tomography is $N$ independent measurements described by the same POVM $\{\Lambda_i\}_{i=1}^k$, on identically prepared quantum systems, each described by a density operator $\rho$. The measurements yield a data set $\delta = \{n_1, n_2, \ldots, n_k\}$, where $n_i$ is the number of measurement outcomes corresponding to $\Lambda_i$, and $\sum_i n_i = N$. The probability $\text{Pr}(\delta|\rho)$ of observing the data set $\delta$ given the density operator $\rho$ is

$$\text{Pr}(\delta|\rho) = C_\delta \prod_{i=1}^k p_i^{n_i},$$

where $C_\delta = N/(n_1! \ldots (n_k!)$ is a normalization constant that only depends on $\delta$.

The linear inversion estimator $\hat{\rho}_{\text{inv}}$ is a simple method for estimating a system’s density operator $\rho$ when the measurement data is related to $\rho$ by a set of invertible linear equations. An example of this general strategy is the measurement scheme discussed above when the POVM $\{\Lambda_i\}_{i=1}^{d^2}$ forms a basis of $L(H)$. The dual basis $\{\bar{\Lambda}_i\}$ is defined by

$$\langle \bar{\Lambda}_i, \Lambda_j \rangle = \delta_{ij}, \quad i, j \in \{1, \ldots, d^2\},$$

where $\langle \rho, \sigma \rangle = \text{Tr}(\rho^1 \sigma)$ is the Frobenius inner product. We estimate $p_i$ by $\hat{p}_i = n_i/N$. The set of invertible linear equations,

$$\text{Tr}(\hat{\rho}_{\text{inv}} \Lambda_i) = \hat{p}_i, \quad i = 1, \ldots, d^2,$$

can be solved to obtain the linear inversion estimator

$$\hat{\rho}_{\text{inv}} = \sum_{i=1}^{d^2} \hat{p}_i \bar{\Lambda}_i.$$
3 Bayesian Estimators

3.1 Bayes Rule

For Bayesian estimators one chooses a prior probability for the estimand, and a model that relates observed data to the estimand, using Bayes rule (discussed below), the prior is updated to obtain a posterior probability, the latter is used to construct point or set estimates. We will be focusing on point estimates.

Let \( \Pr(\rho) \) be a prior probability measure on \( S \), it represents the belief or uncertainty about the quantum system prior to the measurement. For quantum state tomography, measurements are performed on many copies of a quantum system to generate a discrete data set \( \delta \). A model \( \Pr(\delta|\rho) \) is chosen, it relates \( \delta \) to \( \rho \) (see eq. (2) for an example) and represents the probability of obtaining the data given the quantum state. In the literature for a fixed \( \delta \), \( \Pr(\delta|\rho) \) is often viewed as a non-negative function of \( \rho \) called the likelihood function (for more examples see \([9, 10]\)). The posterior probability measure \( \Pr(\rho|\delta) \) on \( S \) given the data is obtained from Bayes rule

\[
\Pr(\rho|\delta) = k \Pr(\delta|\rho) \Pr(\rho),
\]

where \( k \), which depends on \( \delta \) and not on \( \rho \), is a normalization constant.

As \( \rho \) varies continuously, a useful way to express \( \Pr(\rho) \) is by making \( \rho \) a smooth one to one function \( \rho(x) \) of a collection of real variable \( X \), and introducing a non-negative prior probability density \( p(\rho) \) such that

\[
\Pr(\rho \in A) = \int_A p(\rho(x))dx,
\]

where \( A \subseteq S \) is some subset of density operators, and \( A \subseteq X \) the corresponding subset of parameters. The posterior probability density

\[
p(\rho|\delta) = k \Pr(\delta|\rho)p(\rho),
\]

uses the same parametrization \( X \) as before, and is related to the measure \( \Pr(\rho|\delta) \) in a manner similar to (7). Note that for a given probability measure \( \Pr(\rho) \) the density \( p(\rho) \) depends on the choice of parametrization \( \rho(x) \) conversely, if \( p(\rho) \) is held fixed, a different parametrization will lead to a different measure \( \Pr(\rho) \). The same holds true for the relationship of \( \Pr(\rho|\delta) \) and \( p(\rho|\delta) \).

The Bayes mean estimator \( \hat{\rho}_{BME} \) is the expectation of \( \rho \) in the posterior probability measure \( \Pr(\rho|\delta) \), and can be written using the density \( p(\rho|\delta) \) as

\[
\hat{\rho}_{BME} = \int \rho(x)p(\rho(x)|\delta)dx.
\]

While the terms in the integrand depend on the parametrization \( X \) used for \( \rho(x) \), the integral itself is independent of the parametrization as long as \( \Pr(\rho|\delta) \) is held fixed. However if \( p(\rho|\delta) \) is held fixed, changing the parametrization may change \( \Pr(\rho|\delta) \) (see comments following (8)) and alter \( \hat{\rho}_{BME} \). Evaluating the integral (9) numerically can be cumbersome as for \( n \) qubits \( X \) consists of \( 2^n - 1 \) variables.

3.2 MAP estimate

The maximum a posteriori (MAP) estimate \( \hat{\rho}_{MAP} \), is the density operator \( \rho \) for which the posterior probability density is maximum. Its advantage is that in many cases it can be easily computed. When the data set is large, one expects for a suitable prior that \( \hat{\rho}_{BME} \) and \( \hat{\rho}_{MAP} \) are close.

Maximizing \( p(\rho|\delta) \) is equivalent to maximizing \( \log p(\rho|\delta) \), and since \( k \) is independent of \( \rho \), it follows from (8) that

\[
\hat{\rho}_{MAP} = \arg\max_{\rho \in S} [\log \Pr(\delta|\rho) + \log p(\rho)].
\]

Notice that for a fixed \( \Pr(\rho) \), and thus \( \Pr(\rho|\delta) \) as given by (8), \( \hat{\rho}_{MAP} \) will depend on the parametrization \( X \) for \( \rho(x) \). See the comments above in connection with (7). Conversely, if \( p(\rho) \) and thus \( p(\rho|\delta) \), (8), is held fixed, changing the parametrization may change \( \Pr(\rho) \) and \( \Pr(\rho|\delta) \) without altering \( \hat{\rho}_{MAP} \).

It is often the case in quantum tomography that \( \log \Pr(\delta|\rho) \) is concave in \( \rho \) (see eq. (2) for an example). If in addition, as is the case for a number of priors (see the discussion below), \( \log p(\rho) \) is a concave function of \( \rho \) then the same is true for the objective function on the right hand side of (10), and \( \hat{\rho}_{MAP} \) can be efficiently computed using tools of convex optimization.
If one knows that \( \rho \) belongs to a discrete set of possibilities, the above discussion is modified in an obvious way: \( \hat{\rho}_{\text{BME}} \) is the weighted average of finitely many density operators computed with respect to the left hand side in (10), and \( \hat{\rho}_{\text{MAP}} \) is the density operator for which the left hand side in (10) is maximum.

The MLE, HMLE, and MLME estimators can be viewed as MAP estimators using suitable prior probability densities. The maximum likelihood estimate (MLE)

\[
\hat{\rho}_{\text{MLE}} = \arg \max_{\rho \in \mathcal{S}} \log \Pr(\delta|\rho),
\]

coincides with \( \hat{\rho}_{\text{MAP}} \) in (10) when the prior probability density \( p(\rho) \) is independent of \( \rho \). While MLE and the MAP estimate are the same for this special choice of prior, it must be noted that: the former is the density operator for which the data is most likely, and the latter is the most probable density operator given the data and the prior probability density.

The hedged maximum likelihood estimate (HMLE)

\[
\hat{\rho}_{\text{HMLE}} = \arg \max_{\rho \in \mathcal{S}} [\log \Pr(\delta|\rho) + \log \det(\rho)^\beta], \quad \beta > 0,
\]

is of the MAP form, where the prior \( p(\rho) \propto \det(\rho)^\beta \) is called the hedging function; it guarantees a full rank estimate. When \( \beta \) is an integer, this prior probability density can be viewed as a special case of induced measures (see eq. (3.5) in [11]) obtained by choosing an ancillary system of dimension \( k \), defining the Haar measure on a \( d \times k \) dimensional Hilbert space, then tracing out the ancillary system to induce a distribution on the space of \( d \) dimensional density operators. The function \( \log \det(\rho)^\beta \) is concave in \( \rho \) for \( \beta > 0 \), and when \( \log \Pr(\delta|\rho) \) is concave the HMLE can be efficiently computed.

The maximum likelihood maximum entropy (MLME) estimate

\[
\hat{\rho}_{\text{MLME}} = \arg \max_{\rho \in \mathcal{S}} [\log \Pr(\delta|\rho) + \lambda S(\rho)], \quad \lambda \geq 0,
\]

is a MAP estimate with a prior which is exponential in the von-Neumann entropy \( S(\rho) = -\text{Tr}(\rho \log \rho) \). Since \( S(\rho) \) is concave in \( \rho \), when \( \Pr(\delta|\rho) \) is concave the MLME can be efficiently computed. Other possible advantages of the MLME estimator have been discussed in [3].

### 3.3 MAP and Quantum Process Tomography

Let \( \mathcal{H}_a, \mathcal{H}_a' \) and \( \mathcal{H}_b \) be finite dimensional Hilbert spaces with dimensions \( d_a = d_{a'} = d \) and \( d_b \), respectively. Let \( \mathcal{N} : \mathcal{L}(\mathcal{H}_a') \rightarrow \mathcal{L}(\mathcal{H}_b) \) be a quantum channel, and \( \mathcal{I}_a : \mathcal{L}(\mathcal{H}_a) \rightarrow \mathcal{L}(\mathcal{H}_a) \) be the identity map on operators. Let \( \{|a_i\}\) and \( \{|a_i'\}\) be orthonormal basis of \( \mathcal{H}_a \) and \( \mathcal{H}_a' \) respectively, and \( |\phi\rangle = \sum_i |a_i\rangle|a_i'\rangle/\sqrt{d} \) be a maximally entangled bipartite state. The channel \( \mathcal{N} \) can be completely characterized by a bipartite quantum state, sometimes called the Choi matrix or the dynamical operator

\[
\Upsilon = (\mathcal{I} \otimes \mathcal{N})|\phi\rangle\langle\phi|, \quad \Upsilon \in \mathcal{L}(\mathcal{H}_{ab}).
\]

The channel \( \mathcal{N} \) is completely positive, iff the operator \( \Upsilon \) is positive semi-definite (12), see [13] for a diagrammatic proof), and \( \mathcal{N} \) is trace preserving iff

\[
\text{Tr}_b(\Upsilon) = \mathbb{1}_a/d,
\]

(15)

where \( \text{Tr}_b \) is the partial trace over \( \mathcal{H}_b \), and \( \mathbb{1}_a \) is the identity operator on \( \mathcal{H}_a \). For any \( A \in \mathcal{L}(\mathcal{H}_a') \), one can show that

\[
\mathcal{N}(A) = d \text{Tr}_a[(A^T \otimes \mathbb{1}_b)\Upsilon],
\]

(16)

where \( A^T \) denotes the transpose of \( A \) in the \( \{|a'_i\}\) basis. Using eq. (16), the action of the map \( \mathcal{N} \) can be computed using the Choi matrix \( \Upsilon \) alone. Equation (14) gives a one to one correspondence between \( \mathcal{M}_{ab} \): the convex set of quantum channels mapping \( \mathcal{L}(\mathcal{H}_a') \) to \( \mathcal{L}(\mathcal{H}_b) \), and \( \mathcal{T}_{ab} \): the convex set of density operators in \( \mathcal{L}(\mathcal{H}_{ab}) \) with partial trace on \( \mathcal{H}_b \) equaling \( \mathbb{1}_a/d \). This correspondence can be used to construct a MAP estimator for a quantum channel as follows.

Suppose measurements are performed with an aim to characterize the quantum channel \( \mathcal{N} \) (see [14, 15] for examples) and data \( \delta \) is collected. As in in Sec. 3.2 let \( p(\Upsilon) \) be a prior probability density on \( \mathcal{T}_{ab} \), and \( \Pr(\delta|\Upsilon) \) the probability of obtaining \( \delta \) given \( \Upsilon \). The MAP estimate for the Choi matrix

\[
\hat{\Upsilon}_{\text{MAP}} = \arg \max_{\Upsilon \in \mathcal{T}_{ab}} [\log \Pr(\delta|\Upsilon) + \log p(\Upsilon)],
\]

(17)
becomes a MAP estimate $\hat{N}_{\text{MAP}}$ for the quantum channel when $\hat{Y}_{\text{MAP}}$ is inserted in (16). When the objective function on the right hand side of (17) is concave in $Y$, $\hat{Y}_{\text{MAP}}$ can be efficiently computed using tools of convex optimization.

4 Modelling Noise

Noise is present in any experimental setup. If its effect upon a tomographic measurement can be modelled by assuming a known noisy channel $N$ (whose parameters have been determined by previous calibration measurements) preceding the final measurement as in Figure 1, Bayesian estimators for $\rho$ can be obtained by replacing $\Pr(\delta|\rho)$ with $\Pr(\delta|N(\rho))$ in (6) and (8). This results in a modification of the MAP estimate in (10)

$$\hat{\rho}_{\text{MAP}} = \arg\max_{\rho \in S} \left[ \log p(\delta|N(\rho)) + \log p(\rho) \right],$$

(18)

If the measurement in Figure 1 is replaced with a different quantum circuit, then the MAP estimate for the input to this circuit is

$$\hat{\sigma}_{\text{MAP}} = N(\hat{\rho}_{\text{MAP}}).$$

(19)

The above construction is quite general. There is no restriction on $N$, the form of $\Pr(\delta|\rho)$, $p(\rho)$, or the size of the quantum system $d$. Because $N$ is a linear map, if $\log \Pr(\delta|\rho)$ is concave in $\rho$ so is $\log \Pr(\delta|N(\rho))$. Thus when the tools of convex optimization allow an efficient calculation of the MAP estimate in (10) the same will be true of (18). Since the MLE, MLME, and the HMLE are special cases of MAP, they can be adapted in the noisy setting.

Note that the Gaussian noise models considered in [9,10] are quite different: they are not based on a noisy channel, but instead on a special form of $\Pr(\delta|\rho)$.

5 Conclusion

The maximum a posteriori (MAP) estimation framework for quantum state and process tomography introduced here combines a number of previous quantum state estimators, in particular the maximum likelihood, hedged maximum likelihood, and the maximum likelihood-maximum entropy estimator, in a single framework using Bayesian methodology and different priors. In several cases of interest to quantum state tomography the MAP estimator becomes a convex optimization problem which should be numerically more tractable than the Bayes mean estimator. As shown in Sec. 3.3, using the Choi-Jamiołkowski isomorphism MAP estimation of quantum states can be easily extended to quantum channels, and the extension is expected to have similar advantages as the MAP estimate for quantum states. When the experimental noise can be represented by a known noisy channel of the kind considered in Sec. 4 the MAP estimator can be modified to take it into account in a way which in many cases should again allow for an efficient numerical solution.

Acknowledgements

I am indebted to Robert B. Griffiths for valuable comments.

References

[1] Z. Hradil. Quantum-state estimation. Phys. Rev. A, 55:R1561–R1564, Mar 1997.

[2] Robin Blume-Kohout. Hedged maximum likelihood quantum state estimation. Phys. Rev. Lett., 105:200504, Nov 2010.
[3] Yong Siah Teo, Huangjun Zhu, Berthold-Georg Englert, Jaroslav Řeháček, and Zden ěk Hradil. Quantum-state reconstruction by maximizing likelihood and entropy. *Phys. Rev. Lett.*, 107:020404, Jul 2011.

[4] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, Oct 2010.

[5] Robin Blume-Kohout. Robust error bars for quantum tomography. pages 1–7, Feb 2012.

[6] Christopher Ferrie and Robin Blume-Kohout. Minimax Quantum Tomography: Estimators and Relative Entropy Bounds. *Physical Review Letters*, 116(9):090407, mar 2016.

[7] Matthias Christandl and Renato Renner. Reliable quantum state tomography. *Physical Review Letters*, 109(12), 2012.

[8] Robin Blume-Kohout. Optimal, reliable estimation of quantum states. *New Journal of Physics*, 12(4):043034, Apr 2010.

[9] John A. Smolin, Jay M. Gambetta, and Graeme Smith. Efficient Method for Computing the Maximum-Likelihood Quantum State from Measurements with Additive Gaussian Noise. *Physical Review Letters*, 108(7):070502, Feb 2012.

[10] Harpreet Singh, Arvind, and Kavita Dorai. Constructing valid density matrices on an NMR quantum information processor via maximum likelihood estimation. *Physics Letters A*, 380(38):3051–3056, Sep 2016.

[11] Karol Zyczkowski and Hans-Jürgen Sommers. Induced measures in the space of mixed quantum states. *Journal of Physics A: Mathematical and General*, 34(35):7111, 2001.

[12] Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10(3):285 – 290, 1975.

[13] Patrick J. Coles, Li Yu, Vlad Gheorghiu, and Robert B. Griffiths. Information-theoretic treatment of tripartite systems and quantum channels. *Phys. Rev. A*, 83:062338, Jun 2011.

[14] Debbie W. Leung. Choi’s proof as a recipe for quantum process tomography. *Journal of Mathematical Physics*, 44(2):528–533, 2003.

[15] Isaac L.Chuang and M. A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*, 44(11-12):2455–2467, 1997.