View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by ScholarBank@NUS

Error Regions for Properties of The Quantum State

Li Xikun

2016

Error Regions for Properties of The Quantum State

LI XIKUN

M.Sc.,Graduate University of Chinese Academy of Sciences

A THESIS SUBMITTED FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

Centre for Quantum Technologies

National University of Singapore

2016

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Li Xikun

20 February 2016

Dedicated to my wife, parents

and teachers...

Acknowledgments

First of all, I thank my supervisor Prof. Berthold-Georg Englert for his intellectual support over the years. I am deeply grateful for his continued guidance. I also thank Prof. Valerio Scarani and Dr. Wang Qinghai for serving on my thesis advisory committee. Many thanks to Asst/Prof. Ng Hui Khoon, Shang Jiangwei for their advice. I benefited a great deal from numerous stimulating discussions with them. Thanks go as well to Arun Sehrawat and Philippe Raynal for their collaborations and useful discussions. I also want to thank Hu Yuxin, Dai Jibo, Len Yink Loong, Seah Yi-Lin and Ye Luyao for being great colleagues.

Special thanks to my friends Zhu Huangjun and Gao Meng for their various support and help. I also thank my friends You Jiabin, Mei Feng, Yao Penghui and Ved Prakash for the sharing. I would like to acknowledge the financial support from Centre for Quantum Technologies, and thank the administrative staff in CQT for the numerous timely help.

Finally, many thanks to my wife and parents for their continuous love, support and understanding.

Contents

Summary

This thesis mainly studies the method for constructing error intervals for properties of the quantum state.

As a complement to point estimators for the quantum state estimation, region (interval for one dimension) estimators are proposed to supplement the error regions to the point estimator. These proposals, however, are ad hoc because they usually rely on having a lot of data, or consider all the possible data that haven't been observed. In [1], a method is provided for systematically constructing optimal error regions for quantum state estimation from the data actually observed. After identifying the prior probability as the size of a region, two types of optimal error regions– maximum-likelihood regions and smallest credible regions–are reported which are the bounded-likelihood regions that comprise all states with likelihood exceeding a threshold value.

As a generalization of the above scenario for reporting optimal error regions for quantum state estimation, we propose a systematic method for constructing error intervals for a property of state directly from the experimental data. Usually, we are not interested in the full details of the quantum state, but rather care about some parameters or a few properties of the state. Moreover, it is much more difficult to estimate a high-dimensional quantum state. Therefore, a direct estimate of the properties of interest is more practical than the estimate of the whole quantum state. Analogous to error regions for quantum state estimation, the optimal error intervals are characterized by finding the constant likelihood values conditional on the property of state. For illustration, we identify the optimal error intervals for fidelity (with respect to certain target states) and purity of single-qubit states, as well as the CHSH quantity for two-qubit states.

List of Tables

List of Figures

List of Symbols[1]

$C_{\mathcal{R}}$	Credibility of region $\mathcal{R}$	26
c_λ	Credibility of $\mathcal{R}_\lambda$	30
d	Dimension of the Hilbert space	*
$(\mathrm{d}\rho)$	Volume element of the infinitesimal vicinity of the state ρ	*
f_k	Frequencies	14
$\mathcal{H}$	Hilbert space	*
$h(\rho)$	Hedging function	17
I	Identity operator	*
$L(D)$	Prior likelihood	25
$L(D\|\widehat{p}_{\mathrm{ML}})$	The maximal value of $L(D\|p)$	22
$L(D\|p)$	Point likelihood function	22
$L(D\|\mathcal{R})$	Region likelihood	26
N	The number of states used in state tomography	*
n_k	Clicks of the kth detector	*
p_k	Probabilities	*
P_k	Projective measurement	12
$\widehat{\mathcal{R}}_{\mathrm{ML}}$	Maximum-likelihood region (MLR)	*
$\partial\widehat{\mathcal{R}}_{\mathrm{ML}}$	The boundary of $\widehat{\mathcal{R}}_{\mathrm{ML}}$	28
$\mathcal{R}_\lambda$	Bounded-likelihood region (BLR)	29
S_0	The size of reconstruction space $\mathcal{R}_0$	24
$S_{\mathcal{R}}$	Size of region $\mathcal{R}$	24
s_λ	Size of $\mathcal{R}_\lambda$	29
$\mathrm{tr}\{\cdot\}$	Trace of an ordinary operator	*

[1]The page number where a symbol is defined is listed at the rightmost column. When the definition is general, the page number is given as *.

$w_{\text{basic}}(p)$	Basic constraints for permissible probabilities	23
$w_{\text{cstr}}(p)$	Constraints for permissible probabilities	22
$w_{\text{qu}}(p)$	Quantum constraints for permissible probabilities	23
β	Hedging parameter	17
$\delta(\cdot)$	Dirac's delta function	*
$\eta(\cdot)$	Heaviside's unit step function	*
λ	Fraction of $L(D\|\widehat{p}_{\text{ML}})$	*
$\chi_\lambda(p)$	Characteristic function	29
w_0	Prior density	24
ρ	A generic quantum state	*
$\widehat{\rho}$	An estimator of ρ	15
$\widehat{\rho}_{\text{ML}}$	Maximum-likelihood estimator	*
$\sigma_x, \sigma_y, \sigma_z$	Pauli operators	*
Π_k	Measurement outcome	*
Λ_k	Reconstruction operator	15

List of Abbreviations

AAQPT	Ancilla-assisted quantum process tomography
BLI	Bounded-likelihood interval
BLR	Bounded-likelihood region
BM	Bayesian mean
BME	Bayesian mean estimator
CHSH	Clauser-Horne-Shimony-Holt
CPU	Central processing unit
CRLB	Cramér-Rao lower bound
DCQD	Direct characterization of quantum dynamics
DSPE	Direct state-property estimation
EAQPT	Entanglement-assisted quantum process tomography
FIM	Fisher information matrix
HMLE	Hedged maximum-likelihood estimation
HS	Hilbert-Schmidt
IC	Informationally complete
i.i.d	Independently and identically
ILS	Iso-likelihood surface
ISPE	Indirect state-property estimation
MC	Monte Carlo
ME	Maximum entropy
ML	Maximum likelihood
MLE	Maximum-likelihood estimation
MLI	Maximum-likelihood interval
MLR	Maximum-likelihood region
MSE	Mean square error

PM	Projective measurement
POM	Probability operator measurement
POVM	Positive operator valued measure
PPBS	Partially polarizing beam splitter
QDT	Quantum detector tomography
QET	Quantum estimation theory
QSE	Quantum state estimation
QST	Quantum state tomography
RIBF	Regularized incomplete beta function
SCI	Smallest credible interval
SCR	Smallest credible region
SPE	State-property estimation
SQPT	Standard quantum process tomography
TAT	Trine antitrine
TM	Tetrahedron measurement
vNM	von Neumann measurement

Chapter 1

Introduction

Quantum estimation theory (QET), as the name suggests, is a combination of statistical estimation methods and quantum theory. In classical estimation theory, methods are developed to estimate the parameters of the probability density function from a set of data [2]. In analogy to that, QET seeks strategies for estimating parameters in a reliable, optimal and efficient way. The main departure of QET from its classical counterpart is that the underlying theory of QET is quantum mechanics, which includes complementarity principle and uncertainty relation that results in many challenges during the process of QET.

As an important application of quantum estimation theory, quantum tomography [3], like its classical counterpart which produces a three-dimensional image by a series of two-dimensional projections along different directions, is a procedure of inferring the quantum state (density operator), the process matrix, or the quantum measurement, corresponding to quantum state tomography [4–11], quantum process tomography [12–18] and quantum detector tomography [19–22], respectively[1]. These three types of tomography technique are interrelated: given the input and output, we are able to reconstruct the object to be estimated by using various estimation approaches. Quantum tomography is central to many quantum information tasks, such as quantum computation, quantum communication and quantum cryptography, because quantum information processing

[1] In many literatures, quantum measurement tomography, which is a synonym for quantum detector tomography, is adopted as the procedure of referring the measurement that is performed on the quantum system.

includes three components: preparation of quantum state, transiting quantum states through some channel and performing the measurements on quantum states by detectors. All these tasks require reliable quantum tomography at various stages of quantum information processing.

Take quantum state tomography (also called quantum state estimation (QSE)) for example. Given the data from the generalized measurements, known as probability operator measurements (POMs), quantum state estimation can infer the quantum state that provides complete information of the system. In classical physics, the state is described by the set of canonical coordinates. It is always possible, at least in principle, to *fully* reconstructed the state by performing measurement on it. In quantum mechanics, however, it is not possible, even in principle to perfectly determine the quantum state.

Previous works on quantum tomography have yielded a *single* estimated quantum state, process matrix or quantum measurement. Mathematically, this single object is a *point estimator* which serves as a 'best guess' for the unknown (actual) one. In contrast, the region estimator is a set of possible values of the unknown objects to be estimated. Such regions are typically confidence region in the paradigm of frequentist inference or credible region in the case of Bayesian inference. In this thesis, we focus on the credible regions for quantum estimation.

In another aspect, we are interested in a few properties of the state, rather then the full details of the quantum systems. In addition, it is difficult to do quantum state estimation for high-dimensional quantum system. Then, a direct estimate of the few properties of interest is more desirable than the estimate of the whole quantum state. As linear and nonlinear functions of the quantum state, these properties of the quantum state include its fidelity with a target state, purity, and measure of entanglement. Note, however, that even if we were

able to obtain a best guess for the quantum state, the values of the properties in that state may not be the best guess for them. Therefore, state-property estimation (SPE) is needed to be supplemented with quantum state estimation.

Chapter 2 provides an overview of quantum tomography from the theoretical view. We start with a historical survey of developments of the classical estimation theory. Then we provide a brief introduction and historical background of quantum estimation theory. Several ingredients, which is basic block in quantum tomography, are introduced, such as quantum states, measurements, point estimators for quantum state estimation. Prominent methods are listed as examples for quantum tomography techniques. We first present linear inversion which is the simplest method, as well as the well-known maximum-likelihood estimation, followed by hedged maximum-likelihood estimation.

In Chapter 3, we introduce several notions, which are needed for Chapter 4, of optimal error regions for quantum state estimation. Among them, the most important notion is that the size of a region is its prior probability. Also important are the concept of maximum-likelihood region and smallest credible region which are both bounded-likelihood regions.

Chapter 4 presents the theoretical analysis of the state-property estimation. Analogous to Chapter 3, the size and credibility of a range of state-property are identified. Again we find that the optimal error intervals for property of the state are the bounded-likelihood intervals, where the likelihood is F-likelihood, the likelihood conditional on the property F, in the context of SPE. To narrow the choices of approximation function for F-likelihood, two important properties of the F-likelihood are described.

In Chapter 5 we investigate the numerical procedures to determine the F-likelihood. We first employ the Monte Carlo integration to calculate the size and credibility. Since the finite size of the sample, these numerical integrals with the

fluctuations require smoothing approximation fitting. We list 5 candidates for the fitting function, and choose the incomplete beta functions to perform this job. In addition, an iteration algorithm is introduced to verify the quality of the approximation. As illustrations, fidelity (with respect to target state) and purity of single-qubit state, as well as the CHSH quantity of two-qubit state, are studied.

We close with a short conclusion and outlook in Chapter 6.

Chapter 2

Quantum Estimation Theory

2.1 Classical estimation theory

A good quantum estimation strategy involves two important elements: the measurement scheme and the data processing protocols. Given the measurement data, the optimization of data processing is basically a subject of classical estimation theory[1]. Put differently, classical estimation theory always sheds light on the study of quantum estimation theory.

Classical estimation theory is the scientific discipline which deals with the problems caused by discrepancies in different measurements of the same object. These discrepancies, or *errors*, are generally regarded as random in independent observations of the same quantity. To reduce the effect of errors in determining the values of the quantity of interest, scientist need to investigate the problem of defining a suitable estimator. A good estimator should be a well-defined function of the observations whose values serve as a suitable approximation of the quantity of interest. The random nature of the errors implies that the definition of an estimator involves considerations of a relatively arbitrary nature. That means there is generally much freedom in choosing the estimators. Many desirable estimator characteristics need to be considered, such as bias, consistency and efficiency. In the following, we will briefly review the historical development of

[1]Notice that additional constraints imposed by quantum mechanics should be taken into account, such as the positivity of the density matrix.

the classical estimation theory.

The problem of estimating parameters from observation data can be traced from antiquity. It is well known that astronomical studies have stimulated the development of estimation theory even to the present day. From about 300 B.C., Babylonian astronomers dealt with this problem. Since the age of telescope, the motion of the planetary and cometary bodies is observed by using telescopic measurements. These motions can be characterized by six parameters. The problem of estimation is inferring the values of these parameters from the measurement data. The earliest and simplest approach to this problem is the *arithmetic mean*. That is, the estimator of a parameter from independent measurement is taken to be their arithmetic average [23]. It was not until the 18th century the field of estimation theory began to be broadend by other considerations. This broadening coincided with the development of probability theory. In 1722 Roger Cotes considered a planar estimation problem in which he considered the combination of four measurements of a point and assigned weights to each measurement. Then he asserted that the center of gravity of the points is the most probable value of the point. In 1777 D. Bernoulli proposed that the uniform distribution may not describe the measurement errors and that other descriptions may be more appropriate. In addition, Bernoulli developed an estimation procedure known as the method of maximum likelihood in present-day literature. Bernoulli claimed that the estimate should be chosen to correspond to the point which maximizes the likelihood function.

Another result of fundamental importance to estimation theory is due to Thomas Bayes. In 1763 'An essay towards solving a problem in the doctrine of chances' was published. In this essay, Bayes introduced the notion of priori density, posteriori density and Bayes' rule [24]. However, the dispute arises from the existence of a priori distribution, although the mathematical validity is

unquestionable.

Astronomical studies provided the stimulus for invention and development of the method of least-squares. It is believed that Legendre and Gauss developed this method independently. The principle of least-squares is that, given the measurement data, the estimate of the parameters should be chosen to minimize the sum of the squares of the errors. Following Gauss, Laplace and Markov obtained similar statement of least-square principle [25, 26]. A fundamental break from the least-squares method came in a series of papers by Fisher. As an undergraduate at Cambridge, he reinvented the method of maximum likelihood [27]. Then, Fisher introduced the concepts of sufficient statistics, efficient and consistent estimates, which further stimulate the development of classical estimation theory [28, 29].

The methods we mentioned above have their applications in the context of quantum estimation theory (See Sec. 2.5). There are many methods and algorithms, which can be utilized in quantum estimation theory, to be explored in the classical estimation theory.

2.2 Quantum state estimation

Quantum state estimation is of fundamental importance for the quantum information theory. It is also the basic ingredient for quantum process estimation and quantum detector estimation. In the following, we provide a historical survey of the development of quantum estimation theory in general and quantum state estimation in particular; see [3] for a review.

It was Pauli who first asked whether the wave function of a quantum state can be determined by the distribution of position and momentum [30]. This problem was first addressed in 1957 by Fano [31] who introduced the concept of the *quorum*, the set of observables sufficient for completely estimating the

quantum system. The early works [32–36] considered the state reconstruction for one-dimensional spin and spinless particles as examples. However, it is difficult to devise measurable observables other than position, momentum and energy. Vogel and Risken [4] show that by using homodyne measurement, the collection of probability distributions obtained is the Radon transform of the Wigner function. Then the density matrix of that quantum system can be determined based on their correspondence with Wigner functions. The first experiment of quantum tomography was realized using balanced homodyne detector by Smithey *et al.* [5]. This pioneering experiment marked the birth of optical homodyne tomography. Since that time, many new techniques have been proposed, see [37] for a overview.

The development of quantum tomography theory was further stimulated by the advance of experimental techniques and the emergence of quantum information science. Analogous to its classical counterpart–arithmetic mean, linear inversion is the first and easiest estimation method. However, it suffers from two defects: non-positivity and choice ambiguity. By using means of Jaynes principle of maximum entropy (ME) [38,39], Bužek *et al.* [40,41] proposed a method which selects the most objective estimator among all possible candidates, to address the problem of reconstructing quantum states from informationally incomplete measurements. Meanwhile, Hradil developed an efficient algorithm for computing the maximum-likelihood estimation (MLE), which provides one semi-definite positive estimator from the data of detection. Many other reconstruction methods are provided, including hedged maximum-likelihood estimator (HMLE) [42], the Bayesian mean (BM) estimator [2,43–48], and minimax mean estimator [49–53].

As a generalization of quantum state tomography, quantum process tomography and quantum detector tomography are proposed for characterizing un-

known quantum processes and quantum measurements, respectively. Analogous to quantum state tomography, quantum process tomography and quantum detector tomography are central to many tasks of quantum information process.

As the first proposal of quantum process tomography, standard quantum process tomography (SQPT) was introduced by Chuang and Nielsen [13] , as well as by Poyatos *et al.* [14]. In order to characterize a quantum process, a set of reference states is prepared and then subjected to a given (unknown) quantum process. Using the quantum state estimation to identify the resulting quantum states, the quantum processes can be reconstructed. Another technique, as an alternative to SQPT, ancilla-assisted quantum process tomography (AAQPT) was proposed by Leung [54, 55], as well as by D'Ariano and Presti [56], and realized in experiment [57, 58]. AAQPT requires a *single* preparation and tomography of a two-qubit state by introducing an extra *ancilla* qubit. Entanglement assisted quantum process tomography (EAQPT), as a special case, describes the situation when the ancilla is maximally entangled with the system to be characterized. Both of these two methods are known as indirect methods, since they use quantum state estimation to reconstruct the unknown process. In contrast with the above methods, direct characterization of quantum dynamics (DCQD) was devised by Mohseni and Lidar [59, 60] and followed by experiment in which the dynamics of a photon qubit is determined [61] and that of nuclear spins in the solid state [62]. DCQD has many advantages since it does not need quantum state tomography. Therefore, it requires much fewer experiment settings. See Ref. [63] for overview of these three techniques of quantum process tomography.

On the other hand, quantum detector tomography is a procedure to characterize experimental apparatus. To ensure the positivity of probabilities of outcome, POMs are used to represent the detector measurements. Given the input

(known) states and the outcomes of the detectors, the POMs can be estimated by techniques of quantum tomography [19, 21]. For instance, like its quantum state estimation counterpart, we can invert the Born rule to find the corresponding POMs. Since linear inversion sometimes cannot provide physical POMs, the MLE method, which guarantees the positivity of the POMs, are preferred for quantum detector tomography [20]. Lundeen *et al.* realize the experiments of quantum detector tomography which reconstructs two types of detectors: Commercial avalanche photodiode and home-made single photon detector [22].

2.3 Quantum states

The state of a quantum system encodes all the information about the system. Mathematically, the *state space* of the quantum system is a complex vector space, *i.e.*, Hilbert space. The *pure* state is described by a normalized complex vector $|\phi\rangle$, known as *ket* by Dirac's definition:

$$|\phi\rangle = (c_1, c_2, \cdots, c_n)^{\mathrm{T}}, \ \sum_{k=1}^{n} |c_k|^2 = 1, \tag{2.1}$$

where T is transpose of the vector, and n is the dimension of the state space. The generic states are mixed, which are represented by a positive semidefinite matrix with unit trace:

$$\rho \geq 0, \ \mathrm{tr}\{\rho\} = 1. \tag{2.2}$$

As a special case, the density matrix of a pure state can be expressed as

$$\rho_{\mathrm{pure}} = |\phi\rangle\langle\phi|. \tag{2.3}$$

Accordingly, pure states are density operators of rank one, i.e., a projector onto the subspace spaned by state vector. The purity of a quantum state is the trace

of the square of the density matrix: $P \equiv \mathrm{tr}\{\rho^2\}$. Therefore, the purity of pure states is 1, and that of mixed states is less than 1:

$$\mathrm{tr}\{\rho_{\mathrm{pure}}^2\} = 1,\ \mathrm{tr}\{\rho_{\mathrm{mix}}^2\} < 1. \tag{2.4}$$

Composite systems, composed from simple systems, reveal much more interesting phenomena than simple systems. The most famous one is the quantum correlation, known as quantum entanglement (see Ref. [64] for a review), which was emphasized by Einstein [65] and Schrödinger [66]. Quantum entanglement plays an important role in quantum computation [67], as well as many tasks in quantum information processing [64], such as quantum key distribution [68], quantum teleportation [69] and superdense coding [70]. The state space of the composite systems is the tensor product of the Hilbert space of the simple systems. Take a bipartite composite system for instance. Suppose the Hilbert spaces of Alice and Bob are $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, then the Hilbert space $\mathcal{H}$ of the composite system is $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$. Tracing the composite systems over the Hilbert space of subsystem B, we get the *reduced* state of the subsystem A:

$$\rho_A = \mathrm{tr}_B\{\rho_{AB}\}, \tag{2.5}$$

and vice versa.

2.4 Measurements

2.4.1 Projective measurements

Projective measurements (PM) (also called von Neumann measurements (vNM)) are hermitian projectors with eigenvalue 1. They form a complete orthogonal

basis:

$$\sum_k P_k = I,\ P_k P_j = \delta_{kj} P_k, \tag{2.6}$$

where I is the identity matrix. Given the quantum state ρ before measurement, the probability p_k of the outcome k is given by Born rule

$$p_k = \mathrm{tr}\{P_k \rho P_k^\dagger\} = \mathrm{tr}\{P_k \rho\}. \tag{2.7}$$

If the outcome k obtained, then the quantum state after the measurement is represented by state operator

$$\rho' = \frac{P_k \rho P_k^\dagger}{\mathrm{tr}\{P_k \rho P_k^\dagger\}}. \tag{2.8}$$

Since repeated projective measurements always yield the same outcome as the first one, no additional information about the original quantum system is available.

2.4.2 Non-projective measurements

Non-projective measurements generalize the projective ones. Such set of measurement operators only require the completeness condition

$$\sum_k M_k^\dagger M_k = I. \tag{2.9}$$

The probability p_k of the outcome k and the quantum state after the measurement is

$$p_k = \mathrm{tr}\{M_k \rho M_k^\dagger\},\ \rho' = \frac{M_k \rho M_k^\dagger}{\mathrm{tr}\{M_k \rho M_k^\dagger\}}, \tag{2.10}$$

respectively.

If we are only interested in the outcome statistics, we effectively describe the

measurement by a set of positive operators $\Pi_k = M_k M_k^\dagger$ which sum to identity. Literally, such measurement are called as probability operator measurement (POM), or positive operator valued measurement (POVM) which is a synonym for POM. It is known that the element of POM is positive semidefinite matrix, without necessarily being orthogonal

$$\Pi_k > 0, \sum_k \Pi_k = 1, \Pi_k \Pi_j \neq \delta_{kj} \Pi_k. \tag{2.11}$$

The POMs Π_k may be identified with the measurement outcomes. According to Neumark's dilation theorem [71], any set of POMs can be realized by applying a projective measurement on a larger system. Without the need for worrying about the detailed realization of the measurement, POM allow us to focus on the system under study. Compared with projective measurements, POMs are generally easier to handle because of their nicer mathematical structure. For instance, POM reserve the convexity, *i.e.*, any convex combination of POMs is still a POM. This property is important for constructing sophisticated POMs from simple ones.

A measurement is informationally complete (IC) if any quantum state is completely determined by the outcome of measurements [72]. In a d-dimensional Hilbert space, an IC measurement consists of at least d^2 outcomes. While for a *minimal* IC measurement, it consists of d^2 outcomes exactly. For example, for a two-dimensional system, i.e., qubit state, one possible IC measurement is the set of projective measurements

$$\Pi_{k\pm} = \frac{1}{6}(I \pm \sigma_k), \text{ for } k = x, y, z, \tag{2.12}$$

where I is the two-dimensional identity matrix, and the σ_ks are the Pauli matrixes. For this IC measurement, there are six outcomes. However, a minimal IC

measurement requires only 4 outcomes for qubit state. Such minimal IC measurement can be constructed by linearly combining the identity matrix and Pauli matrixes (see Ref. [11]).

2.5 Point estimators for quantum state estimation

After the data D are collected from the measurement apparatus, the remaining task is to provide a best guess, i.e., point estimator, for the quantum state based on the data. In classical statistics, many methods are proposed for constructing point estimators. In the context of quantum state estimation, those classical methods indeed shed light on inferring the quantum state. However, additional requirements from quantum mechanics, such as the positivity of the density matrix, should be taken into account. In the following, we list several common methods of point estimators for quantum state estimation, and discuss the pros and cons of each method. See Refs. [3, 37] for a more detailed review.

2.5.1 Linear inversion

Analogous to its classical counterpart–arithmetic mean, linear inversion is one of the simplest point estimators. It was first studied by Fano [31], followed by many other researchers [32, 34–36]. Recently, linear inversion is re-used for constructing error bars which are confidence intervals [73].

Given measurements specified by the POM with unit sum $\sum_k \Pi_k = 1$, N copies of identical and independently prepared quantum systems are measured with the detector clicks $\sum_k n_k = N$. Then the relative frequency of the kth outcome is $f_k = n_k/N$. In linear inversion, the point estimator is the state which

matches the observed frequencies:

$$\mathrm{tr}\{\rho\Pi_k\} = f_k \quad \text{for } k = 1, 2, \cdots, K\,. \tag{2.13}$$

For IC measurement, at most one $\hat{\rho}$ exists as the solution. Moreover, for the minimal IC measurement, only one unique solution exists. In such case, a unique dual basis Λ_k exists with the property $\mathrm{tr}\{\Pi_k\Lambda_k\} = \delta_{kl}$. Then it is easy to derive the estimator for the state $\hat{\rho} = \sum_k f_k\Lambda_k$. However, since the number of copies is finite, the statistical noise associated with the relative frequencies is inevitable. Therefore, for a generic IC measurement, the system of equations in Eq. (2.13) might be incompatible. In other words, such estimators which are compatible with the frequencies do not exist.

The main benefit we have from linear inversion is its simplicity. Its drawbacks are also obvious. First, the estimator given by linear inversion is sometimes not positive semidefinite. That means the state is not a legitimate one. This phenomena is common when the actual states are of high purity or the sample size is small. Second, if the measurement is informationally overcomplete, then the solution is not unique and thus some arbitrariness is introduced during the choice of the reconstruction operators. To solve these problems, MLE is proposed as as alternative method which is discussed in the next section.

2.5.2 Maximum-likelihood estimation

As we mentioned in Sec. 2.1, the principle of MLE was reinvented by Fisher and becomes a basic ingredient in classical statistics [28]. Moreover, the principle of MLE is extensively applied in the context of quantum state estimation [3, 37, 74]. Another application of MLE is the detection [42] and characterization of entanglement [75]. Instead of searching for the quantum state which matches frequencies obtained from the data of outcomes, MLE seeks the quantum state

which maximizes the likelihood function (see Sec. 3.2.2) defined as

$$L(D|p) = p_1^{n_1} p_2^{n_2} \cdots p_K^{n_K}, \tag{2.14}$$

where $D = \{n_1, n_2, \cdots, n_K\}$ is the clicks of detectors and $p_k = \mathrm{tr}\{\rho\Pi_k\}$ is the probability of obtaining the outcome k given the true state ρ. In practice, it is more convenient to work with the *log-likelihood function* which is concave

$$\ln L(D|p) = \sum_{k=1}^{K} n_k \ln p_k = N \sum_{k=1}^{K} f_k \ln p_k, \tag{2.15}$$

If state exists that matches the observed frequencies, *i.e.*, satisfying Eq.(2.13), then this state is also a MLE.

Generally, however, it is difficult to analytically find a closed formula for MLE. Hradil [74] proposed an algorithm that can compute efficiently MLE. Because of the concavity of log-likelihood function and convexity of the state space, the search for the MLE turns to be a convex optimization problem, which can be solved based on the principle of steepest ascent. See Ref. [74] for more information. The MLE can also suffer from the problem of arbitrariness, if the measurement is not IC. To solve this problem, Teo *et al* provide a solution [76] based on the ML principle [28] and the ME principle [38,39]. It is proposed that an efficient algorithm for computing the most objective estimator, which is the state having the highest von Neumann entropy among all the states with the maximum value of the likelihood function.

The MLE is widely used because the estimator is always positive semidefinite. However, there is a drawback of the ML estimation, i.e., the zero eigenvalue problem. When the true state ρ has a high purity, the estimator $\hat{\rho}$ is not often full rank. The zero eigenvalues imply unrealistic confidence over the outcomes of some potential measurements, which is undesirable for some quantum infor-

mation processing tasks, such as data compression and cryptography [48]. To solve this zero-eigenvalue problem, the hedged maximum-likelihood estimation is proposed which is the topic of next section.

2.5.3 Hedged maximum-likelihood estimation

Analogous to its classical counterpart which is known as "add β" rule [77], the hedged maximum-likelihood estimation (HMLE) is proposed by Blume-Kohout to solve the zero-eigenvalue problem [42]. In HMLE, the likelihood function $L(D|p)$ is multiplied by a hedging function

$$h(\rho) = \det(\rho)^{\beta}, \tag{2.16}$$

where the value of the hedging parameter β is between 0 and 1. Thus the functional to be maximized is $L(D|p)h(\rho)$, rather than $L(D|p)$. The estimator of HMLE is guaranteed to be full rank. Because of the concavity of the function $\ln[L(D|p)h(\rho)]$, a similar algorithm can be applied for computing efficiently the HLME estimator. Although HMLE can solve the problem of zero-eigenvalue, the lack of universal criterion for choosing the hedging function makes it difficult to determine the value of β. Even if we may rely on the prior knowledge, such kind of choice is made on an ad hoc basis.

Chapter 3

Error Regions for Quantum State Estimation

3.1 Introduction

Quantum state estimation is important to many tasks of quantum information processing which makes use of quantum sources, such as characterizing the source of quantum-information carriers, verifying the properties of a quantum channel and monitoring the transmission line used for quantum key distribution. In general, a collection of independently and identically (i.i.d) prepared copies of quantum systems (photons, electrons, ions, etc.) are measured one-by-one, *i.e.*, adapting individual measurements, by an apparatus which can be mathematically described by a probability-operator measurement (POM)[1]. The POM consists of a number of outcomes which register individual information carriers, and the data are the observed sequence of detector clicks. Mathematically, the quantum state to be estimated, the *actual state*, is described by a statistical operator. And the measurement data are used to infer an *estimator*—another state which approximates the actual state well. Many strategies are studied for finding such an estimator (see Sec. 2.5). Because of its efficient algorithm, maximum likelihood estimators (MLEs) developed by Hradil, Řeháček, and their collaborators

[1] Here we do not consider collective measurements which may be more efficient for extracting information than individual ones.

are employed in real experiments for analysing the data (reviewed in [3],; see also [78] and Sec. 2.5.2). Given the data, the MLE is the state for which the data are more likely than for any other state. Meanwhile, we can choose other methods to reconstruct a point estimator. (see Sec. 2.5)

However, no matter which method we choose, the data have statistical noise. Therefore, point estimators cannot identify the actual state exactly. Put differently, we need to supplement the point estimator with error bars (*error regions* for higher-dimensional problems). Several methods are proposed for attaching a vicinity of states to a point estimator. However, these methods are often ad hoc in nature. They usually rely on obtaining a lot of data [79, 80], involve data resampling [81] or consider all possible data that might have been observed [82,83]. In contrast to them, we propose a method which systematically constructs error regions from the data *actually* observed, and the data are not necessarily many.

In this chapter, [2] we propose *maximum-likelihood regions* (MLRs) and *smallest credible regions* (SCRs), which are optimal error region for quantum state estimations. In Sec. 3.2, we introduce the concepts of the reconstruction space and constraints on the permissable probabilities, as well as the notion of the size of a region, which is of importance to MLR and SCR. Then we solve the optimization problems that identify the MLRs and SCRs and find their solutions in Sec. 3.3.

[2] This chapter is based on Refs. [1] and [84], I sincerely acknowledge Dr. Shang and other authors of Ref. [1].

3.2 Ingredients of optimal error regions

3.2.1 Reconstruction space, likelihood and constraints

In a real experiment, the data we collect from the detectors are detection events, or clicks. The Born rule links these clicks to the theoretical predictions which quantum mechanics could tell us. Suppose the POM has K outcomes $\Pi_1, \Pi_2, \ldots, \Pi_K$, which are positive Hilbert space operators with unit sum. The probability p_k for the kth outcome is

$$p_k = \mathrm{tr}\{\Pi_k \rho\} = \langle \Pi_k \rangle \ . \tag{3.1}$$

The positivity of ρ and its unit trace ensure that the p_ks satisfy the basic constraints

$$p_k \geq 0 \,, \quad \sum_{k=1}^{K} p_k = 1 \,. \tag{3.2}$$

For the chosen POM, however, the Born rule implies further constraints on p_ks. For instance, consider a single-qubit state measured by four-outcome tetrahedron POM with probabilities [11]

$$\begin{aligned} p_1 &= \frac{1}{4}(1-z) \,, \quad p_2 = \frac{1}{4}\left(1 + \sqrt{\frac{8}{9}}y + \frac{1}{3}z\right), \\ \left.\begin{matrix} p_3 \\ p_4 \end{matrix}\right\} &= \frac{1}{4}\left(1 \pm \sqrt{\frac{2}{3}}x - \sqrt{\frac{2}{9}}y + \frac{1}{3}z\right). \end{aligned} \tag{3.3}$$

where $\boldsymbol{r} = (x, y, z)$ is the Bloch vector with $x = \langle \sigma_x \rangle$, $y = \langle \sigma_y \rangle$, and $z = \langle \sigma_z \rangle$. These tetrahedron probabilities are further constrained by

$$0 \leq p_k \leq \frac{1}{2} \,, \quad \frac{1}{4} \leq \sum_{k=1}^{4} p_k^2 \leq \frac{1}{3} \,, \tag{3.4}$$

which results from the fact that the length of the Bloch vector does not exceed one: $|r| \leq 1$. Generically, for high-dimensional systems (e.g., two-qubit states) measured by many-outcome POMs, the constraints imposed by the Born rule are not easy to state explicitly as conditions on the probability p_k.

A probability $p = (p_1, p_2, \ldots, p_k)$ is called *physical* or *permissible* if it satisfies all the constraints, *i.e.*, Eq. (3.1) holds. All the permissible probabilities constitute the *probability space*. Given ρ, the probability p is uniquely determined by Eq. (3.1). The converse, however, may not be true if the POM is *not* informationally complete (IC). Therefore, if there are several ρs for the same p, we pick up *one* representative, and thus have a one-to-one mapping $\rho \leftrightarrow p$. These representative ρs constitute the *reconstruction space* $\mathcal{R}_0$. This one-to-one mapping identifies ρ with p, and regions in the reconstruction space with the corresponding regions in the probability space. We should notice that, however, it is not always possible to find a convex reconstruction space, though the probability space is convex. Thus, it is convenient to work in the probability space.

The point likelihood $L(D|p)$ is the probability of obtaining D if p is the true state (see Eq. (2.14)).[3] Note that point likelihood $L(D|p)$ has one important property: $\log(L(D|p))$ is concave. That means $L(D|p)$ has a unique (local) maximum value: $L(D|p)$ takes one its largest value for the maximum-likelihood estimator (MLE)

$$L(D|p)_{\max} \equiv \max_{p \in \mathcal{R}_0} L(D|p) \equiv L(D|\widehat{p}_{\mathrm{ML}})\,. \tag{3.5}$$

The constraints on permissible probabilities $w_{\mathrm{cstr}}(p)$ is the product of ba-

[3]Since there is an one-to-one mapping between (permissible) probability space and reconstruction space, *i.e.*,$\rho \leftrightarrow p$, we call p the state, and other way round. Readers should not be confused and can distinguish them according to context.

sic constraints $w_{\text{basic}}(p)$ and quantum constraints $w_{\text{qu}}(p)$ imposed by quantum mechanics

$$w_{\text{cstr}}(p) = w_{\text{basic}}(p) w_{\text{qu}}(p), \tag{3.6}$$

where $w_{\text{bacis}}(p)$, which is always contained in $w_{\text{cstr}}(p)$, is the product of Heaviside's unit step function $\eta(\cdot)$ and Dirac's delta function $\delta(\cdot)$

$$w_{\text{basic}}(p) \equiv \eta(p_1)\eta(p_2)\cdots\eta(p_K)\delta\left(\sum_{k=1}^{K} p_k - 1\right), \tag{3.7}$$

while the quantum constraints w_{qu} contains the product of step functions which are imposed by quantum mechanics. For the tetrahedron POM (see Eqs. (3.3) and (3.4)), the quantum constraint is

$$w_{\text{qu}}(p) = \eta\left(\frac{1}{2} - p_1\right)\eta\left(\frac{1}{2} - p_2\right)\cdots\eta\left(\frac{1}{2} - p_4\right)\eta\left(\frac{1}{3} - \sum_{k=1}^{4} p_k^2\right). \tag{3.8}$$

3.2.2 Prior probability as the size of a region

To construct a *region* for quantum state estimation, a important concept has to be clarified: what is the *size* of a region? Consistent with the argument in [85], we measure the size of a region by its prior probability. In other words, the size of a region expresses the relative importance of that regions of states. We denote by $(\mathrm{d}\rho)$ the volume element of the infinitesimal vicinity of the state ρ in $\mathcal{R}_0$

$$(\mathrm{d}\rho) = (\mathrm{d}p) w_0(p), \tag{3.9}$$

where $(\mathrm{d}p)$ is the volume element of the probability space incorporating the constraints

$$(\mathrm{d}p) = \mathrm{d}p_1 \mathrm{d}p_1 \ldots \mathrm{d}p_K \, w_{\text{cstr}}(p), \tag{3.10}$$

and $w_0(p)$ is the (unnormalized) prior density. Note that we exclude the cases of improper priors where the prior density $(\mathrm{d}\rho)$ cannot be normarlized. Two specific priors are used that are *non-informative* . The first one is the *primitive prior*,

$$w_{\text{primitive}}(p) = 1, \tag{3.11}$$

such that the density is uniform in p over the probability space. The second one is the *Jeffreys prior*

$$w_{\text{Jeffreys}}(p) = \frac{1}{\sqrt{p_1 p_2 \cdots p_K}}, \tag{3.12}$$

which is a popular choice of unprejudiced prior. Both primitive prior and Jeffreys prior are non-informative prior in which no external prior information is available. For a discussion of various aspects of prior selection, see Appendix A in [1].

Given the prior density $w_0(p)$, the size $S_{\mathcal{R}}$ of a region $\mathcal{R}$ is

$$S_{\mathcal{R}} = \frac{\int_{\mathcal{R}}(\mathrm{d}\rho)}{\int_{\mathcal{R}_0}(\mathrm{d}\rho)} = \frac{\int_{\mathcal{R}}(\mathrm{d}p)w_0(p)}{\int_{\mathcal{R}_0}(\mathrm{d}p)w_0(p)} \equiv \frac{1}{S_0}\int_{\mathcal{R}}(\mathrm{d}p)w_0(p)\,, \tag{3.13}$$

where the normalization factor $S_0 \equiv \int_{\mathcal{R}_0}(\mathrm{d}p)w_0(p)$ is the size of reconstruction space $\mathcal{R}_0$[4]. From Eq. (3.13), it is clear that $S_{\mathcal{R}}$ does not depend on the parametrization used for numerical representation of $(\mathrm{d}\rho)$. The primary parametrization is in terms of the probabilities p.

As an example, consider a harmonic oscillator with its infinite-dimensional state space. Suppose the POM has two outcomes with p_1 to be the probability that oscillator is in its ground state, and $p_2 = 1 - p_1$. Its probability space is the same as that of a tossed coin. The prior density $w_0(p)$ selects the line segment with $0 \le p_1 = 1 - p_2 \le 1$ in the $p_1 p_2$ plane. First, we choose the primitive prior $(\mathrm{d}\rho) = (\mathrm{d}p)$, such that the subsegment with $a \le p_1 \le b$ has size $b - a$. Then, for

[4] $S_0 = 1$ is implied by Eq. (3.13)

the Jeffreys prior,

$$(\mathrm{d}\rho) = \frac{1}{\pi}(\mathrm{d}p) w_{\mathrm{Jeffreys}}(p) = (\mathrm{d}p)\frac{1}{\pi\sqrt{p_1 p_2}}, \tag{3.14}$$

we have the size $\frac{2}{\pi}[\sin^{-1}(\sqrt{b}) - \sin^{-1}(\sqrt{a})]$ for the same subsegment.

3.2.3 Region likelihood and credibility

The joint probability of finding the state ρ in the region $\mathcal{R}$ and obtaining the data D is

$$\mathrm{prob}(D \wedge \mathcal{R}) = \frac{1}{S_0}\int_{\mathcal{R}} (\mathrm{d}\rho)\, L(D|p)\,. \tag{3.15}$$

For $\mathcal{R} = \mathcal{R}_0$, the *prior likelihood* $L(D)$ is

$$L(D) \equiv \mathrm{prob}(D \wedge \mathcal{R}_0) = \frac{1}{S_0}\int_{\mathcal{R}_0} (\mathrm{d}\rho)\, L(D|p)\,. \tag{3.16}$$

For all the possible data D that can be observed, we have

$$\begin{aligned} \sum_D L(D|p) &= \sum_{n_1,\ldots,n_K} \frac{N!\,\delta_{N,n_1+n_2+\cdots+n_K}}{n_1!\,n_2!\,\cdots\,n_K!} p_1^{n_1} p_2^{n_2}\cdots p_K^{n_K} \\ &= (p_1 + p_2 + \cdots + p_K)^N = 1\,, \\ \sum_D L(D) &= \frac{1}{S_0}\int_{\mathcal{R}_0} (\mathrm{d}\rho) = 1\,, \end{aligned} \tag{3.17}$$

where the summation is over all values of the n_ks and the multinomial factor is the number of sequences with the same counts of detector clicks.

Analogous to Bayes' theorem which is

$$P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)\,,$$

The joint probability $\mathrm{prob}(D \wedge \mathcal{R})$ can be factorized into two different ways

$$\mathrm{prob}(D \wedge \mathcal{R}) = L(D|\mathcal{R}) S_{\mathcal{R}} = C_{\mathcal{R}}(D) L(D)\,, \tag{3.18}$$

and identify the *region likelihood* $L(D|\mathcal{R})$

$$L(D|\mathcal{R}) = \frac{\mathrm{prob}(D \wedge \mathcal{R})}{S_{\mathcal{R}}} = \frac{\int_{\mathcal{R}} (\mathrm{d}\rho) L(D|p)}{\int_{\mathcal{R}} (\mathrm{d}\rho)}\,, \tag{3.19}$$

and the *credibility* $C_{\mathcal{R}}(D)$

$$C_{\mathcal{R}}(D) = \frac{\mathrm{prob}(D \wedge \mathcal{R})}{L(D)} = \frac{\int_{\mathcal{R}} (\mathrm{d}\rho) L(D|p)}{\int_{\mathcal{R}_0} (\mathrm{d}\rho) L(D|p)}\,. \tag{3.20}$$

The region likelihood $L(D|\mathcal{R})$, analogous to point likelihood $L(D|p)$, is the probability of obtaining the data D if the state is in the region $\mathcal{R}$. The credibility $C_{\mathcal{R}}(D)$, which is the posterior probability of the region, is the probability that the actual state is in the region $\mathcal{R}$ if the data D have been obtained.

3.3 Optimal error regions

3.3.1 Maximum-likelihood regions and smallest credible regions

Equation (3.19) implies two possible candidates for the optimal error regions. The first one is the maximum-likelihood regions (MLRs), i.e, the regions which have the largest likelihood. Specifically, for the given size s of the region, we maximize the region likelihood $L(D|\mathcal{R})$

$$\max_{\mathcal{R} \subseteq \mathcal{R}_0} L(D|\mathcal{R}) = L(D|\widehat{\mathcal{R}}_{\mathrm{ML}}) \quad \text{with } S_{\mathcal{R}} = s\,. \tag{3.21}$$

An unconstrained maximization of $L(D|\mathcal{R})$ is not meaningful, because it gives the limiting region that consists of nothing but the point $\widehat{\rho}_{\text{ML}}$, which maximizes the point likelihood $L(D|p)$ (see Eq. (3.5)). Or we get the whole reconstruction space $\widehat{\mathcal{R}}_{\text{ML}} = \mathcal{R}_0$ without any constraints. The MLR $\widehat{\mathcal{R}}_{\text{ML}}$ desired is a function of the data D and the size s. Fixing the size s, we can equivalently maximize the joint probability $\text{prob}(D \wedge \mathcal{R})$ under the size constraint (see Eq. (3.18))

$$\max_{\mathcal{R}} \text{prob}(D \wedge \mathcal{R}) = \text{prob}(D \wedge \widehat{\mathcal{R}}_{\text{ML}}) \quad \text{with } S_{\mathcal{R}} = s\,. \tag{3.22}$$

Based on the second way of factorizing $L(D|\mathcal{R})$, we have another optimal error regions $\widehat{\mathcal{R}}_{\text{sc}}$: smallest credible regions (SCRs). SCR is a region which contains the actual state with high probability. Analogous to MLR, we fix the credibility of the region, and desire that the size of the region to be smallest. For the given data D, the optimization problem for SCR is

$$\min_{\mathcal{R}\subseteq\mathcal{R}_0} S_{\mathcal{R}} = S_{\widehat{\mathcal{R}}_{\text{sc}}} \quad \text{with } C_{\mathcal{R}}(D) = c. \tag{3.23}$$

Since the credibility c is fixed and $L(D)$ is constant (once the data D is given), the joint probability $\text{prob}(D \wedge \mathcal{R})$ is also fixed. That means the minimization problem for SCR can be equivalently stated as

$$\min_{\mathcal{R}\subseteq\mathcal{R}_0} S_{\mathcal{R}} = S_{\widehat{\mathcal{R}}_{\text{sc}}} \quad \text{with fixed } \text{prob}(D \wedge \mathcal{R}). \tag{3.24}$$

Therefore, MLRs and SCRs are dual to each other. For MLRs, we maximize the joint probability for given size; for SCRs, we minimize the size for given joint probability.

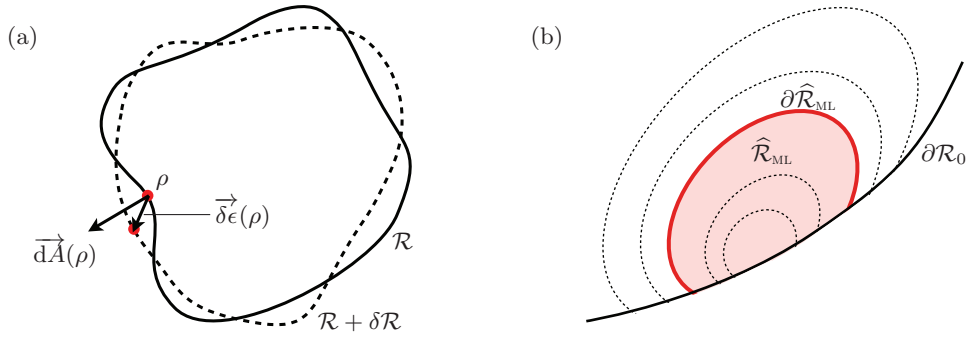

Figure 3.1: (a) Infinitesimal variation of region $\mathcal{R}$. The boundary $\partial\mathcal{R}$ of region $\mathcal{R}$ (solid line) is deformed to the boundary of region $\mathcal{R}+\delta\mathcal{R}$ (dashed line). $\overrightarrow{\mathrm{d}A}(\rho)$ is the vectorial surface element of $\partial\mathcal{R}$ at ρ, and $\overrightarrow{\delta\epsilon}(\rho)$ is the infinitesimal displacement of ρ. (b) Dotted lines indicate ILSs. The boundary $\partial\widehat{\mathcal{R}}_{\mathrm{ML}}$ of $\widehat{\mathcal{R}}_{\mathrm{ML}}$ can contain part of the surface $\partial\mathcal{R}_0$ of the reconstruction space.

3.3.2 Bounded-likelihood regions

In this section, we prove that both MLR and SCR are bounded-likelihood regions (BLRs). Take the maximization problem of MLR for instance[5]. Consider infinitesimal variations of a region $\mathcal{R}$ by deforming its boundary. From Eq. (3.22) we know that the fixed size of region and the maximum property of $\mathrm{prob}(D\wedge\mathcal{R})$ require the the following relations,

$$\begin{aligned}\delta S_{\mathcal{R}} &= \int_{\partial\mathcal{R}} \overrightarrow{\mathrm{d}A}(\rho)\cdot\overrightarrow{\delta\epsilon}(\rho) = 0\,,\\ \delta\mathrm{prob}(D\wedge\mathcal{R}) &= \int_{\partial\mathcal{R}} \overrightarrow{\mathrm{d}A}(\rho)\cdot\overrightarrow{\delta\epsilon}(\rho)\,L(D|p) = 0\,,\end{aligned}$$

which are true for all possible variations about $\mathcal{R}=\widehat{\mathcal{R}}_{\mathrm{ML}}$ (see Fig. 3.1(a)). To satisfy the above relations at the same time, we require that $L(D|p)$ is constant on the boundary $\partial\widehat{\mathcal{R}}_{\mathrm{ML}}$ of $\widehat{\mathcal{R}}_{\mathrm{ML}}$, where $\partial\widehat{\mathcal{R}}_{\mathrm{ML}}$ is an *iso-likelihood surface* (ILS). Concavity of the (log-)likelihood further requires $\widehat{\mathcal{R}}_{\mathrm{ML}}$ to be the interior of this ILS. In another case, $\partial\widehat{\mathcal{R}}_{\mathrm{ML}}$ can also contain part of the boundary of $\mathcal{R}_0$ (see

[5]Note that this conclusion is also true for the minimization problem of SCR, because of the duality between MLR and SCR.

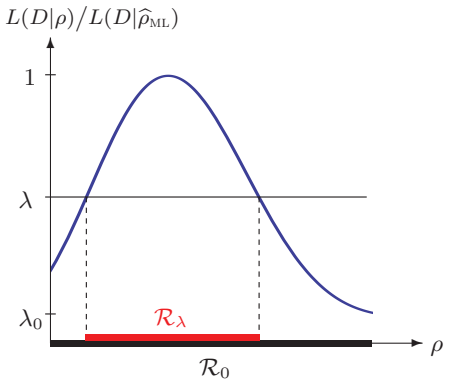

Figure 3.2: Illustration of a BLR: $\mathcal{R}_0$ is the reconstruction space; the region $\mathcal{R}_\lambda$ is a BLR, delineated by the threshold value $\lambda L(D|\widehat{\rho}_{\rm ML})$; λ_0 marks the minimum ratio $L(D|\rho)/L(D|\widehat{\rho}_{\rm ML})$ over $\mathcal{R}_0$.

Fig. 3.1(b)). Put differently, only the part of $\partial\widehat{\mathcal{R}}_{\rm ML}$ inside $\mathcal{R}_0$ is an ILS. In either case, $\widehat{\mathcal{R}}_{\rm ML}$ is a *bounded-likelihood region* (BLR). For the states inside BLR, their point likelihood exceeds a certain threshold. BLRs have appeared previously in standard statistical analysis (see [86] and references therein). It follows that the BLRs are not only the MLRs, they are also the SCRs: each MLR is a SCR, each SCR is a MLR.

The threshold value can be specified as a fraction λ of the maximum value $L(D|\widehat{p}_{\rm ML})$ of the point likelihood; see Fig. 3.2. The characteristic function of a BLR $\mathcal{R}_\lambda$ has the form

$$\chi_\lambda(p) = \eta\Big(L(D|p) - \lambda L(D|\widehat{p}_{\rm ML})\Big)\,, \tag{3.25}$$

and the size of $\mathcal{R}_\lambda$ is

$$s_\lambda = \frac{1}{S_0}\int_{\mathcal{R}_0} (\mathrm{d}\rho)\,\chi_\lambda(p). \tag{3.26}$$

The size of the whole reconstruction space is $s_\lambda = s_0 = 1$ for $\lambda \leq \lambda_0$ with $\lambda_0 \geq 0$ given by $\min_p L(D|p) = \lambda_0\, L(D|\widehat{p}_{\rm ML})$. Increasing λ from λ_0 to 1, the corresponding s_λ decreases monotonically from 1 to 0. As a consequence of the result that MLRs are BLRs, the MLRs always contain the MLE. As the size de-

crease to zero $s \to 0$, the MLR becomes an infinitesimal vicinity of the MLE, and $L(D|\widehat{\mathcal{R}}_{\text{ML}}) \to L(D|\widehat{p}_{\text{ML}})$.

Analogous to s_λ, the credibility of the BLR can be characterized by

$$c_\lambda = \frac{1}{L(D)S_0} \int_{\mathcal{R}_0} (\mathrm{d}\rho)\, \chi_\lambda(p) L(D|p), \tag{3.27}$$

The credibility c specified in (3.23) is obtained for an intermediate value, and the corresponding BLR is the looked-for SCR. It is easy to check that s_λ and c_λ are linked by (see Eqs. (3.26) and (3.27))

$$L(D)\frac{\partial}{\partial\lambda} c_\lambda = L(D|\widehat{p}_{\text{ML}})\, \lambda \frac{\partial}{\partial\lambda} s_\lambda \,. \tag{3.28}$$

As a result, c_λ can be expressed in terms of λ and s_λ[6]

$$c_\lambda = \frac{\lambda s_\lambda + \int_\lambda^1 \mathrm{d}\lambda'\, s_{\lambda'}}{\int_0^1 \mathrm{d}\lambda'\, s_{\lambda'}} \,. \tag{3.29}$$

Relation (3.29) is important for computing the credibility c_λ. From Eq. (3.27) we know that the integral of c_λ requires a well-tailored Monte Carlo integration to handle the point likelihood which is usually sharply peaked. Now we only need to evaluate s_λ according to relation (3.29), such that the numerical effort is substantially reduced. In Fig. 3.3 we demonstrate a simple geometrical meaning of relation (3.29) in terms of areas under the graph of s_λ. Therefore, the BLRs are fully determined by the value of λ, thus can be easily reported. In addition, if one wishes check whether a state is inside the specified error region or not, he/she only need to compute the point likelihood $L(D|p)$ of that state, and compare $L(D|p)$ with $\lambda L(D|\widehat{p}_{\text{ML}})$. See [1] for examples of the BLR of single qubit and two qubits.

[6]For the derivation of Eqs. (3.28) and (3.29), see Appendix A

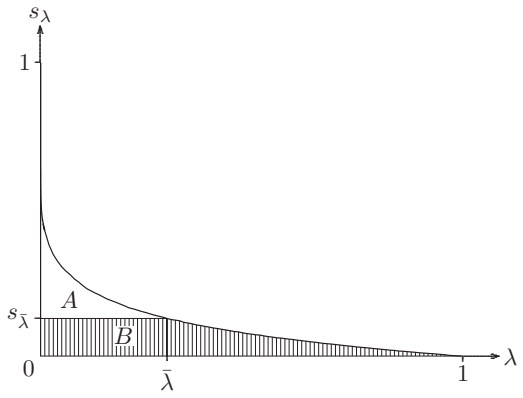

Figure 3.3: Geometrical meaning of the relation (3.29) between the size s_λ and the credibility c_λ. For the chosen value of λ, say $\bar\lambda$, the horizontal line from $(0, s_{\bar\lambda})$ to $(\bar\lambda, s_{\bar\lambda})$ divides the area under the graph of s_λ into the two pieces A and B indicated in the plot. The credibility is the fractional size of the area B, that is: $c_{\bar\lambda} = B/(A+B)$.

3.4 Summary

In this chapter, we describe the reason for why we need error regions for quantum state estimation. As a complement, or improvement, to point estimators, region estimators are proposed in the context of quantum estimation. In contrast to previous work of region estimators, our method only require the data actually observed, and there is no restrictions on the shape of the error regions.

Several notions, such as the reconstruction space and the size of a region, are introduced which are necessary for defining the optimal error regions. Two optimal error regions—MLRs and SCRs—are identified. We prove that both of them are BLRs which can be easily reported as a function of intermediate value λ and point likelihood $L(D|p)$.

Most of the notions and conclusions mentioned in this chapter are reused and re-expressed in the context of state-property estimation which is studied in Chapters 4 and 5. However, the meaning of them here are quite different with the meaning of those notions in Chapters 4, where specific difficulties and challenges are not easy to come by, and thus it is required that a smarter algorithm for

solving the problems.

Chapter 4

Error Regions for State-Property Estimation: Theory

4.1 Introduction

In Chapter 3, we investigate the optimal error regions for QSE. In certain situations, however, we care much about the values of several functions of the state, rather than the full details of the quantum state. For instance, we are interested in the fidelity between the actual state and the target state that are supposed to be emitted from the source. Moreover, it is generally difficult to perform high-dimensional QSE. Therefore, a direct estimate of the few properties of interest is more practical and more desirable than the estimate of the whole quantum state.

We should notice that, even if we can obtain a best guess for the quantum state, i.e., a point estimator for the quantum system, the corresponding values of the properties of interest in that state may not be, and often are not, the best guess for these properties. Then, we need to supplement QSE with SPE — state-property estimation, that is: methods, procedures, and algorithms by which one arrives at an educated guess for the properties of interest directly.

We can regard thess properties, which are the functions of the state, as parameters of the quantum state. Traditionally, however, the term "quantum parameter estimation" is referred to estimation of parameters of the experimental appara-

tus, such as the phase of an interferometer loop [87] and the efficiencies of the detectors [88]. See [88] for a review of various aspects of the combined optimal error regions for the parameters of both kinds. Another branch which is related to the quantum parameter estimation is quantum metrology [89, 90]. Quantum metrology is a study of providing higher resolution and sensitive measurement of the physical parameters, such as distance, acceleration, and frequency, by using quantum mechanics.

Several methods have been proposed for the task of direct estimation of properties of the quantum state [91–98]. A simple quantum network to directly estimate functions of the quantum state, such as purity and spectrum of density operator is presented in [91]. Methods of direct detection and characterization of quantum entanglement are proposed in [92, 93], based on positive maps separability criterion and Renyi entropy, respectively. In [96], the likelihood ratio test is used to draw direct conclusions about entanglement. These methods, however, are restricted to specific functions of the quantum state, or particular settings which are hardly applicable to real tomography experiment. In a different viewpoint, direct and fast methods to estimate the fidelity between the high-dimensional pure state and the actual state are studied in [94, 95, 97]. In these works, protocols using fidelity as a measure are proposed to verify that the state of system is the desired many-body entangled quantum state. However, these methods usually rely on the additional symmetry of the states of interest (e.g. Greenberger-Horne-Zeilinger (GHZ) and W states, etc.) [94, 95], or assume the state is pure [97]. By contrast, our method systematically constructs error intervals for any properties of the quantum state without any restrictions. We notice that Faist and Renner [98] proposes a method, which is analogous to our proposal, for constructing error bars for quantum estimation of a figure of merit, e.g., fidelity, on the notion of confidence region. It is, however, in contrast to our

proposal which adopts Bayesion methodology.

As an extension of the optimal error regions for QSE (see Chapte 3), we proposed in this chapter a systematic method for constructing optimal error intervals for SPE[1]. These optimal error intervals, analogous to those for QSE, are bounded-likelihood intervals (BLIs). For illustration, we only study the situation in which a single function of the state, i.e., a single property. The results can be generalized to the cases in which several properties exists. However, we emphasize that these properties should be estimated jointly, because the constraints should be taken into account .

4.2 Size and credibility of a property range

In Chapter 3, we introduced the ingredients of QSE, such as reconstruction space $\mathcal{R}_0$, the prior density $w_0(p)$, the constraints w_{cstr}, the size of a region $S_{\mathcal{R}}$, and so on. In this chapter, we continue to employ those notions[2].

The property to be estimated can be specified as a function $f(p)$ of the probabilities, with values between 0 and 1,

$$0 \leq f(p) \leq 1\,. \tag{4.1}$$

Of course, the restriction to this convenient range can be easily lifted by rescaling the range. In fact, the property is at first a function $\tilde{f}(\rho)$ of the quantum state ρ. In the context of quantum estimation, ρ is parameterized by probabilities p, such that $f(p) = \tilde{f}\big(\rho(p)\big)$ is the implied function of p.

The iso-F hypersurfaces in probability space and the reconstruction space

[1] This chapter and Chapter 5 are based on Ref. [99], I sincerely acknowledge the contributions from the other authors of Ref. [99]

[2] Note that $S_{\mathcal{R}_0} = \int_{\mathcal{R}_0}(\mathrm{d}\rho) = 1$ is implied in this chapter.

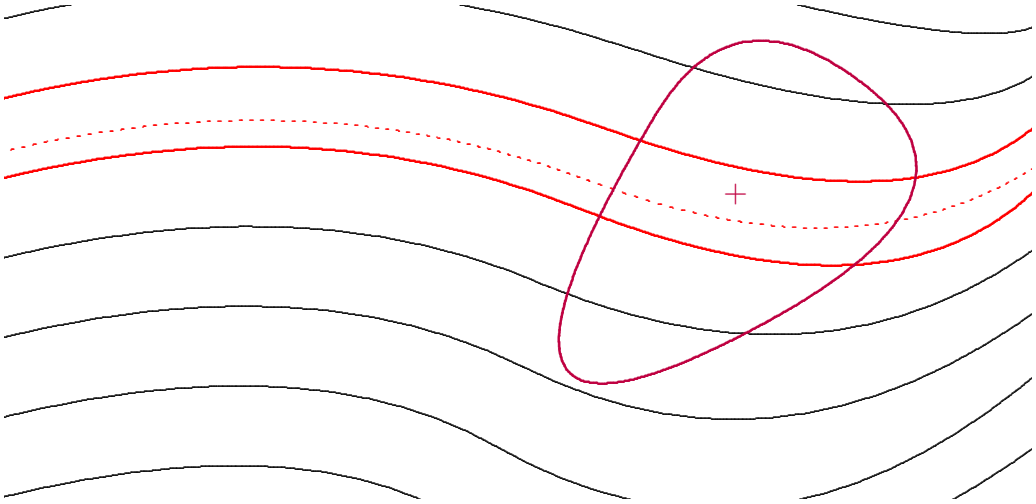

Figure 4.1: Schematic sketch of a sector in the probability space or the reconstruction space. The wave-like lines indicate iso-F hypersurfaces; any two lines mark the boundaries of an F interval, a region specified by a range of F values. The thicker red lines mark the borders of a smallest credible interval (SCI). The dashed red line inside the SCI indicates the hypersurface of the maximum-likelihood estimator $\widehat{F}_{\text{ML}}$. The purple cross marks the maximum-likelihood estimator $\widehat{\rho}_{\text{ML}}$ of the quantum state, with the closed purple curve marking the boundary of the smallest credible region (SCR) with the same credibility as the SCI. Eq. (4.34) states that the purple cross is usually not on the dashed red line, as the plot shows. Note that the SCR contains F values from a larger range than the SCI; see also Fig. 5.4 .

are identified by a given value $F = f(p)$. Therefore, an interval $F_1 \leq f(p) \leq F_2$ corresponds to a region; see Fig. 4.1 for illustration. Such a region has size

$$\begin{aligned} & \int_{\mathcal{R}_0} (\mathrm{d}\rho) \left[\eta\big(F_2 - \tilde{f}(\rho)\big) - \eta\big(F_1 - \tilde{f}(\rho)\big)\right] \\ = & \int (\mathrm{d}p)\, w_0(p) \left[\eta\big(F_2 - f(p)\big) - \eta\big(F_1 - f(p)\big)\right] \\ = & \int (\mathrm{d}p)\, w_0(p) \int_{F_1}^{F_2} \mathrm{d}F\, \delta\big(F - f(p)\big) \end{aligned} \tag{4.2}$$

and credibility

$$\begin{aligned} & \frac{1}{L(D)} \int_{\mathcal{R}_0} (\mathrm{d}\rho)\, L(D|p) \Big[\eta\big(F_2 - \tilde{f}(\rho)\big) - \eta\big(F_1 - \tilde{f}(\rho)\big) \Big] \\ = \; & \frac{1}{L(D)} \int (\mathrm{d}p)\, w_0(p) L(D|p) \int_{F_1}^{F_2} \mathrm{d}F\, \delta\big(F - f(p)\big)\,, \end{aligned} \tag{4.3}$$

The size of prior element $\mathrm{d}F\, W_0(F)$ in F for an infinitesimal slice, $F \leq f(p) \leq F + \mathrm{d}F$, is

$$\mathrm{d}F\, W_0(F) = \int (\mathrm{d}p)\, w_0(p)\, \mathrm{d}F\, \delta\big(F - f(p)\big)\,, \tag{4.4}$$

which implies the prior density of F [3]

$$W_0(F) = \int (\mathrm{d}p)\, w_0(p)\, \delta\big(F - f(p)\big)\,. \tag{4.5}$$

In addition, the credibility Eq. (4.3) tells us the likelihood $L(D|F)$ of the data for given property value F and data D,

$$\frac{1}{L(D)}\, \mathrm{d}F\, W_0(F) L(D|F) = \frac{1}{L(D)} \int (\mathrm{d}p)\, w_0(p) L(D|p)\, \mathrm{d}F\, \delta\big(F - f(p)\big)\,, \tag{4.6}$$

where the likelihood $L(D|F)$ is

$$\begin{aligned} L(D|F) &= \frac{\int (\mathrm{d}p)\, w_0(p) L(D|p)\, \delta\big(F - f(p)\big)}{W_0(F)} \\ &= \frac{W_D(F)}{W_0(F)}\,, \end{aligned} \tag{4.7}$$

here we denote the posterior density by $W_D(F)$, such that we have

$$W_D(F) = W_0(F) L(D|F). \tag{4.8}$$

[3] Notice the difference between that and the prior density in the probability space $w(p)$.

To avoid any potential confusion with the likelihood $L(D|p)$, we shall call $L(D|F)$ the F-likelihood.

The size of a finite interval of F values, or the union of such intervals, denoted by the symbol $\mathcal{I}$, is specified in terms of $W_0(F)$ and $L(D|F)$

$$S_{\mathcal{I}} = \int_{\mathcal{I}} \mathrm{d}F\, W_0(F) \tag{4.9}$$

Analogously, the credibility is

$$C_{\mathcal{I}} = \frac{1}{L(D)} \int_{\mathcal{I}} \mathrm{d}F\, W_0(F) L(D|F) = \frac{1}{L(D)} \int_{\mathcal{I}} \mathrm{d}F\, W_D(F)\,, \tag{4.10}$$

where

$$L(D) = \int_{\mathcal{I}_0} \mathrm{d}F\, W_0(F) L(D|F) = \int_{\mathcal{I}_0} \mathrm{d}F\, W_D(F) \tag{4.11}$$

has the same value as the integral of Eq. (3.16). Denote by $\mathcal{I}_0$ the whole property space of $0 \leq F \leq 1$, such that $S_{\mathcal{I}_0} = C_{\mathcal{I}_0} = 1$ holds.

From another point of view, the F-likelihood $L(D|F)$ can be derived as the derivative of the interval likelihood, the conditional probability

$$\begin{aligned} L(D|\mathcal{I}) &= \frac{\Pr\big(D \wedge \{F \in \mathcal{I}\}\big)}{\Pr(F \in \mathcal{I})} \\ &= \frac{1}{S_{\mathcal{I}}} \int (\mathrm{d}p)\, w_0(p)\, L(D|p) \int_{\mathcal{I}} \mathrm{d}F\, \delta(F - f(p)) \\ &= \frac{1}{S_{\mathcal{I}}} \int (\mathrm{d}p)\, w_0(p) \frac{W_0(F)}{W_0(F)}\, L(D|p) \int_{\mathcal{I}} \mathrm{d}F\, \delta(F - f(p)) \\ &= \frac{1}{S_{\mathcal{I}}} \int_{\mathcal{I}} \mathrm{d}F\, W_0(F) \frac{1}{W_0(F)} \int (\mathrm{d}p)\, w_0(p) L(D|p) \delta(F - f(p))\,. \end{aligned} \tag{4.12}$$

If we now define the F-likelihood by the requirement

$$L(D|\mathcal{I}) = \frac{1}{S_{\mathcal{I}}} \int_{\mathcal{I}} \mathrm{d}F\, W_0(F)\, L(D|F)\,, \tag{4.13}$$

then the expression for $L(D|F)$ in Eq. (4.7) is recovered.

4.3 F-likelihood and its properties

In Chapter 3, the prior density $w_0(p)$ and point likelihood $L(D|F)$ are central for the constructing the optimal error regions for QSE. Given the data D, the BLRs can be easily reported by the characteristic function in which the threshold value is specified by a fraction λ of the maximum values $L(D|\widehat{p}_{\rm ML})$ of the point likelihood $L(D|F)$ (see Eq. (3.25)). This conclusion still holds for constructing the BLIs for SPE. In this section, we study the F-likelihood and its properties.

4.3.1 Free choice of prior

We know that the point likelihood $L(D|p)$ does not depend on the prior density $w_0(p)$. Therefore, the following question arises: Does the F-likelihood $L(D|F)$ depends on the prior density $W_0(F)$? The answer is no.[4]

Assume we have the prior density $w_0(p)$ on the hypersurface where $f(p) = F$,

$$w_0(p)\Big|_{f(p)=F} = W_0(F)u_F(p)\,, \tag{4.14}$$

where the implied prior density $u_F(p)$ assigns relative weights of ps on the iso-F hypersurface. The $u_F(p)$ is normalized

$$\int (\mathrm{d}p)\, u_F(p)\, \delta(F - f(p)) = 1\,, \tag{4.15}$$

[4]Note, however, that the F-likelihood depends on the reference prior $w_{\rm r}(p)$, see below for further information.

which results from the fact that $w_0(p)$ and $W_0(F)$ are also normalized

$$\int (\mathrm{d}p)\, w_0(p) = 1\,, \qquad \int_0^1 \mathrm{d}F\, W_0(F) = 1\,, \tag{4.16}$$

From a different point of view, let us start from the prior density of the property $W_0(F)$. That means we can choose $W_0(F)$ as we like. Now we introduce the prior density $u_{f(p)}$, which can be independently chosen, on the iso-F hypersurfaces. Since F is the coordinate in p-space that is normal to the iso-F hypersurfaces (see Fig. 4.1), the prior density $w_0(p)$ on the whole probability space can be re-defined by these two prior densities

$$w_0(p) = W_0\big(f(p)\big)\, u_{f(p)}(p)\,. \tag{4.17}$$

The restriction to a particular value of $f(p)$ takes us back to Eq. (4.14), as it should.

Given the definition of the the F-likelihood (see Eq. (4.7)) and a prior density of the form Eq. (4.17), the F-likelihood can be specified in the following

$$\begin{aligned} L(D|F) &= \frac{1}{W_0(F)} \int (\mathrm{d}p)\, w_0(p) L(D|p) \delta\big(F - f(p)\big) \\ &= \int (\mathrm{d}p)\, u_F(p) L(D|p) \delta\big(F - f(p)\big) \end{aligned} \tag{4.18}$$

which tells us that $L(D|F)$ does not depend on $W_0(F)$, but is solely determined by $u_F(p)$. In practice, we begin with some reference prior density $w_{\mathrm{r}}(p)$, then the iso-F prior density are chosen to be

$$u_F(p) = \frac{w_{\mathrm{r}}(p)\Big|_{f(p)=F}}{\displaystyle\int (\mathrm{d}p')\, w_{\mathrm{r}}(p')\, \delta\big(F - f(p')\big)} \tag{4.19}$$

that we shall use throughout, such that

$$w_0(p) = \frac{w_{\rm r}(p) W_0(f(p))}{\int (\mathrm{d}p')\, w_{\rm r}(p')\, \delta\big(f(p) - f(p')\big)} \tag{4.20}$$

is the corresponding prior density for the $W_0(F)$ of our liking. As we mentioned in Chapter 3, the normalization of $w_{\rm r}(p)$ is not important; more generally yet, the replacement $w_{\rm r}(p) \to w_{\rm r}(p) g\big(f(p)\big)$ with an arbitrary function $g(F) > 0$ has no effect on the right-hand sides of Eq. (4.19), Eq. (4.20), as well as Eq. (4.21) below.

4.3.2 Two ways to define the F-likelihood

From Eqs. (4.18) and (4.19), the F-likelihood $L(D|F)$ can be expressed in terms of the reference prior $w_{\rm r}(p)$

$$\begin{aligned} L(D|F) &= \frac{\int (\mathrm{d}p)\, w_{\rm r}(p)\, \delta\big(F - f(p)\big) L(D|p)}{\int (\mathrm{d}p)\, w_{\rm r}(p)\, \delta\big(F - f(p)\big)} \\ &= \frac{W_{r,D}(F)}{W_{\rm r,0}(F)}\,, \end{aligned} \tag{4.21}$$

here we denote by $W_{r,D}(F)$ and $W_{\rm r,0}(F)$ numerator and denominator in Eq. (4.21)

$$W_{\rm r,0}(F) = \int (\mathrm{d}p)\, w_{\rm r}(p)\, \delta\big(F - f(p)\big)\,, \tag{4.22}$$

and

$$W_{\rm r,}D(F) = \int (\mathrm{d}p)\, w_{\rm r}(p)\, L(D|p)\, \delta\big(F - f(p)\big)\,, \tag{4.23}$$

respectively[5]. Generally, the difference between $w_0(p)$ and $w_{\mathrm{r}}(p)$ determines the fact that

$$W_0(F) \neq W_{\mathrm{r},0}(F) \;\text{ and }\; W_D(F) \neq W_{\mathrm{r},D}(F). \tag{4.24}$$

Note, the F-likelihood $L(D|F)$ is defined in two different ways — Eqs. (4.7) and (4.21). For Eq. (4.7), the prior density in the integral is $w_0(p)$, while for Eq. (4.21), $w_{\mathrm{r}}(p)$ is employed in the integration. It is easy to verify that these two definition are equivalent to each other: From Eqs. (4.7) and (4.20), we have the F-likelihood.

$$\begin{aligned}
L(D|F) &= \frac{W_D(F)}{W_0(F)} \\
&= \frac{\int (\mathrm{d}p)\, w_0(p)\, \delta\big(F - f(p)\big) L(D|p)}{\int (\mathrm{d}p)\, w_0(p)\, \delta\big(F - f(p)\big)} \\
&= \frac{\int (\mathrm{d}p)\, w_{\mathrm{r}}(p) W_0(f(p))\, \delta\big(F - f(p)\big) L(D|p)}{\int (\mathrm{d}p)\, w_{\mathrm{r}}(p) W_0(f(p))\, \delta\big(F - f(p)\big)} \\
&= \frac{\int (\mathrm{d}p)\, w_{\mathrm{r}}(p)\, W_0(F)\, \delta\big(F - f(p)\big) L(D|p)}{\int (\mathrm{d}p)\, w_{\mathrm{r}}(p)\, W_0(F)\, \delta\big(F - f(p)\big)} \\
&= \frac{W_0(F)}{W_0(F)} \frac{\int (\mathrm{d}p)\, w_{\mathrm{r}}(p)\, \delta\big(F - f(p)\big) L(D|p)}{\int (\mathrm{d}p)\, w_{\mathrm{r}}(p)\, \delta\big(F - f(p)\big)} \\
&= \frac{W_{\mathrm{r},D}(F)}{W_{\mathrm{r},0}(F)}\,,
\end{aligned} \tag{4.25}$$

which is exactly Eq. (4.21).

[5] See Sec. 5.2 for further information.

In practice, it is more convenient to start from the reference priors $w_{\mathrm{r}}(p)$. In the examples in Chapter 5, we employ two different reference priors $w_{\mathrm{r}}(p)$: the primitive prior and the Jeffreys prior (see Eqs. (3.11) and (3.12)).

While the F-likelihood is the same for all $W_0(F)$, it is usually different for different $u_F(p)$s and thus for different $w_{\mathrm{r}}(p)$s. However, analogous to the statements in Chapter 3, the data dominate rather than the priors unless the data are too few. It is because, for sufficiently many data, the point likelihood $L(D|p)$ is narrowly peaked in probability space such that it will be essentially vanishing outside a small region within the iso-F hypersurface. Then it does not matter which reference prior is used.

4.3.3 The properties of the F-likelihood

From the definition of the F-likelihood (see Eq. (4.7)), we know that $L(D|F)$ depends on the reference prior $w_{\mathrm{r}}(p)$ (and $w_0(p)$), the function of state $f(p)$ and point likelihood $L(D|p)$. Obviously, the F-likelihood $L(D|F)$ is different from the point likelihood $L(D|p)$, and it is difficult to determine the properties of $L(D|F)$. However, there are two important properties which should be paid attention to.

Property 1: The F-likelihood $L(D|F)$ is positive.

Since $w_0(p)$ ($w_{\mathrm{r}}(p)$) and $L(D|p)$ are positive, the denominator and numerator in Eq. (4.7) (Eq. (4.21)) are positive, thus it holds for F-likelihood $L(D|F)$.

Property 2: The F-likelihood $L(D|F)$ may not be concave.

It is well known that $\ln L(D|p)$ is concave, while it may not be true for $L(D|F)$. For illustration, we take squared fidelity of one-qubit state as example.

Example: incomplete single-qubit tomography for squared fidelity The three-outcome trine measurement is chosen to be the tomographic measure-

ment [100], with the POM

$$M_1 = \frac{1}{3}\left(\mathbf{1} + \sigma_z\right), \quad \left.\begin{matrix} M_2 \\ M_3 \end{matrix}\right\} = \frac{1}{3}\left(\mathbf{1} \pm \frac{\sqrt{3}}{2}\sigma_x - \frac{1}{2}\sigma_z\right). \tag{4.26}$$

According to the Born rule, the probabilities for the trine outcomes are

$$p_1 = \frac{1}{3}\left(1 + z\right), \quad \left.\begin{matrix} p_2 \\ p_3 \end{matrix}\right\} = \frac{1}{3}\left(1 \pm \frac{\sqrt{3}}{2}x - \frac{1}{2}z\right), \tag{4.27}$$

where $\boldsymbol{r} = (x, z)$ is the two-dimensional Bloch vector with $x = \langle\sigma_x\rangle$ and $z = \langle\sigma_z\rangle$.

The squared fidelity

$$F_{\rm sq} = \mathrm{tr}\{|\sqrt{\rho}\,\sqrt{\rho_{\rm tar}}|\}^2 \tag{4.28}$$

is the measure of overlap between the true state ρ and the target state $\rho_{\rm tar}$. Here we choose the target state $\rho_{\rm tar} = |0\rangle\langle 0|$, such that $F_{\rm sq} = \frac{1}{2}(1+z)$. To derive $L(D|F_{\rm sq})$, we employ the primitive prior $w_{\rm primtive}$. During the integration, we transfer from p-coordinate system to xz coordinate system. Then, the denominator $W_{\rm r,0}(F_{\rm sq})$ in Eq. (4.21) is

$$\begin{aligned} W_{\rm r,0}(F_{\rm sq}) &= \int (\mathrm{d}p)\,\delta\big(F_{\rm sq} - \frac{1}{2}(1+z)\big) \\ &= \frac{1}{\pi}\int_{x^2+y^2\le 1} \mathrm{d}x\;\mathrm{d}y\;\delta\big(F_{\rm sq} - \frac{1}{2}(1+z)\big) \\ &= \frac{2}{\pi}\int_{-2\sqrt{F_{\rm sq}(1-F_{\rm sq})}}^{2\sqrt{F_{\rm sq}(1-F_{\rm sq})}} \mathrm{d}x = \frac{8}{\pi}\sqrt{F_{\rm sq}(1-F_{\rm sq})}. \end{aligned} \tag{4.29}$$

Suppose the clicks of trine outcome is $(n_1, n_2, n_3) = (1, 1, 1)$, then the numerator $W_{{\rm r},D}(F_{\rm sq})$ is

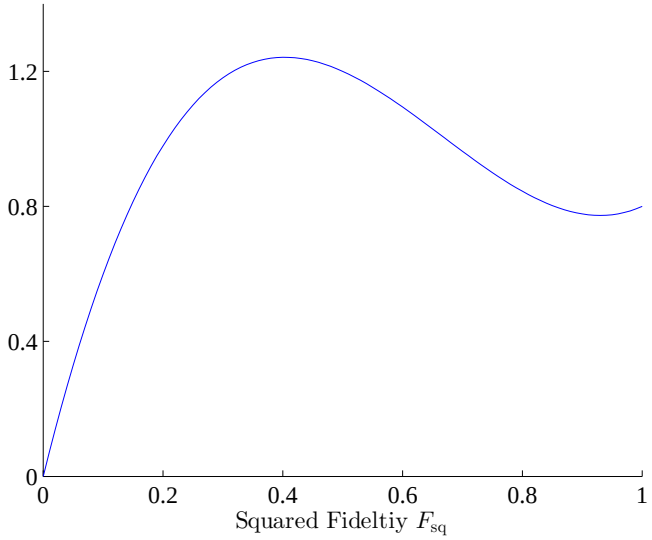

Figure 4.2: The graph of $F_{\rm sq}$-likelihood $L(D|F_{\rm sq})$ (Eq. (4.31)). It is obvious that this cubic polynomial function is neither concave nor convex.

$$\begin{aligned} W_{{\rm r},D}(F_{\rm sq}) &= \int (\mathrm{d}p)\, p_1 p_2 p_3\, \delta\big(F_{\rm sq} - \frac{1}{2}(1+z)\big) \\ &= \frac{32}{5\pi}\sqrt{F_{\rm sq}(1-F_{\rm sq})}\, F_{\rm sq}\, \big(8(1-F_{\rm sq})^2+1\big)\,. \end{aligned} \tag{4.30}$$

Therefore, the F-likelihood, which is the ratio of $W_{{\rm r},D}(F_{\rm sq})$ and $W_{{\rm r},0}(F_{\rm sq})$, is

$$\begin{aligned} L(D|F_{\rm sq}) &= \frac{W_{{\rm r},D}(F_{\rm sq})}{W_{{\rm r},0}(F_{\rm sq})} \\ &= \frac{4}{5}\, F_{\rm sq}\, \big(8(1-F_{\rm sq})^2+1\big)\,. \end{aligned} \tag{4.31}$$

Figure 4.2 shows the F-likelihood of the squared fidelity $F_{\rm sq}$ (with respect to $|0\rangle$) for the primitive prior with the data $D=(1,1,1)$. As a particular example, this cubic polynomial function of $F_{\rm sq}$ (Eq. (4.31)) is neither concave nor convex. In the examples in Chapter 5, however, the F-likelihoods $L(D|F)$ of the properties are unimodal concave functions.

4.4 Optimal error intervals

Even if the F-likelihood $L(D|F)$ may not be concave, the maximum-likelihood estimator $\widehat{F}_{\text{ML}}$ exists

$$\max_F\{L(D|F)\} = L(D|\widehat{F}_{\text{ML}})\,. \tag{4.32}$$

As we mentioned in Sec. 4.1, the value of $\tilde{f}(\rho)$ for point estimators of the quantum state is different from the corresponding estimators for state-property. For example, the maximum-likelihood estimator $\widehat{\rho}_{\text{ML}}$ for the quantum state is

$$\widehat{\rho}_{\text{ML}} = \rho(\widehat{p}_{\text{ML}}) \quad \text{with} \quad \max_p\{L(D|p)\} = L(D|\widehat{p}_{\text{ML}})\,. \tag{4.33}$$

In general, however, the function $\tilde{f}(\rho)$ for MLE of the state is not equal to SPE for the property estimated

$$\tilde{f}(\widehat{\rho}_{\text{ML}}) \neq \widehat{F}_{\text{ML}}\,. \tag{4.34}$$

In full analogy to the OERs for QSE in Chapter 3, the optimal error intervals (OEIs) for property of states $F = \tilde{f}(\rho) = f(p)$ are the bounded-likelihood intervals (BLIs) specified by

$$\mathcal{I}_\lambda = \left\{F \;\middle|\; L(D|F) \geq \lambda L(D|\widehat{F}_{\text{ML}})\right\} \quad \text{with} \quad 0 \leq \lambda \leq 1\,. \tag{4.35}$$

However, we should notice the difference between the OERs for QSE and the OEIs for SPE. Since the point likelihood $L(D|p)$ is concave, the OER for QSE is always connected. However, for the OEIs for SPE, it may not be true, because the F-likelihood may not be concave; see Fig. 4.2. In other words, the OEIs can be a union of finite intervals.

Analogous to the BLR and SCR, the maximum-likelihood interval (MLI) $\widehat{\mathcal{I}}_{\text{ML}}$ is the BLI with the given size $s = S_{\mathcal{I}_\lambda} \equiv s_\lambda$, while the smallest credible interval (SCI) $\widehat{\mathcal{I}}_{\text{SC}}$ is the BLI with the given credibility $c = C_{\mathcal{I}_\lambda} \equiv c_\lambda$. We

have $\mathcal{I}_\lambda \subseteq \mathcal{I}_0$, $s_\lambda \leq s_0 = 1$, and $c_\lambda \leq c_0 = 1$ for $\lambda \leq \lambda_0$, with $\lambda_0 \geq 0$ given by $\min_F L(D|F) = \lambda_0 L(D|\widehat{F}_{\text{ML}})$. As λ increases from λ_0 to 1, s_λ and c_λ decreases monotonically from 1 to 0. Again, we have the link between s_λ and c_λ, which is exactly same as that for BLRs for QSE (see Eq. (3.29)) [6],

$$c_\lambda = \frac{\lambda s_\lambda + \int_\lambda^1 \mathrm{d}\lambda' \, s_{\lambda'}}{\int_0^1 \mathrm{d}\lambda' \, s_{\lambda'}} \,. \tag{4.36}$$

Note, however, that the link between c_λ and s_λ here is not as important as the one in Chapter 3. In Chapter 3, given λ and point likelihood $L(D|p)$, c_λ can be computed without much numerical effort. In the context of SPE, F-likelihood $L(D|F)$ is required to compute c_λ. However, to compute $L(D|F)$, we need to evaluate the numerator $W_{\text{r},D}(F)$ in Eq. (4.21), in which a sophisticated MC integration is required to handle the sharply peaked point likelihood $L(D|p)$.

Once the F-likelihood $L(D|F)$ is given, it is easy to find the MLIs and the SCIs. Compared with MLI, we are more interested in the SCI for the desired credibility c: The actual value of F is contained in this SCI with probability c. Each BLI, with given credibility or size, reports an error bar on $\widehat{F}_{\text{ML}}$ in which is the maximum-likelihood estimator for SPE.

[6] See Appendix A as a reference.

Chapter 5

Error Regions for State-Property Estimation: Numerical

5.1 Introduction

In chapter 4 we introduced several notions: the size element Eq. (4.4), the credibility element Eq. (4.6), and the F-likelihood Eq. (4.7) (or Eq. (4.21)) which are of importance for the construction of error intervals for F. However, we should notice that, in general, these integrals cannot be integrated analytically. Put differently, the integrals involved are usually high-dimensional, such that they can only be computed numerically by Monte Carlo (MC) methods. Moreover, to compute $L(D|F)$, we need to obtain $W_{\mathrm{r},0}(F)$ and $W_{\mathrm{r},D}(F)$ at the first step. The problem is that the numerical values for $W_{\mathrm{r},0}(F)$ and $W_{\mathrm{r},D}(F)$ are zeros for a finite interval, while the analytical values should be positive (but close to zero). Then, as the ratio of $W_{\mathrm{r},D}(F)$ and $W_{\mathrm{r},0}(F)$, the F-likelihood will be infinite, or undefined at that interval. Such a situation should be avoided.

In this chapter, we provide an efficient numerical algorithm that solves the high-dimensional integrals for the size, credibility and F-likelihood. In Sec. 5.2, we describe a iteration algorithm which provides us a proper approximation for $W_{\mathrm{r},0}(F)$ and $W_{\mathrm{r},D}(F)$. We illustrate the matter by a simulated single-qubit experiment in Sec. 5.3, where the fidelity (with respect to target state) and purity are studied as examples. In Sec.5.4 the CHSH quantity of a two-qubit state are

investigated, and the SCIs for fix-vectors and optimized CHSH quantity are reported.

5.2 Numerical procedures

5.2.1 Numerical integrals for the size and credibility

Obviously, the integrands in Eqs. (4.22) and (4.23), in which the delta-function factors exist, are not suitable for a MC integration. Rather, we consider the antiderivatives which contain the step function

$$P_{\mathrm{r},0}(F) = \int (\mathrm{d}p)\, w_{\mathrm{r}}(p)\, \eta\big(F - f(p)\big) \tag{5.1}$$

and

$$P_{\mathrm{r},D}(F) = \frac{1}{L(D)} \int (\mathrm{d}p)\, w_{\mathrm{r}}(p) L(D|p)\, \eta\big(F - f(p)\big)\,. \tag{5.2}$$

These are the prior and posterior contents of the interval $0 \leq f(p) \leq F$ for the reference prior with density $w_{\mathrm{r}}(p)$.

At first, we focus on the denominator $W_{\mathrm{r},0}(F)$ (see Eq. (4.22)). Using the Hamiltonian Monte Carlo sampling described in [101] and [102], we sample the probability space in accordance with the prior $w_{\mathrm{r}}(p)$ and due attention to $w_{\mathrm{cstr}}(p)$ of Eq. (3.6). Since the size of the random sample is finite, the resulting numerical values contain fluctuations. Moreover, certain regions in the probability space cannot be reached, because that the size is finite. Then, the integration values are zero, while the analytical values should not be. Therefore, as we mentioned earlier, we cannot differentiate this numerical approximation of $P_{\mathrm{r},0}(F)$. The solution is: Fit a several-parameter smoothing function to the values produced by the MC integration, and then differentiate this function and so arrive at an approximation $\widetilde{W}_{\mathrm{r},0}(F)$ for $W_{\mathrm{r},0}(F)$.

5.2.2 Candidates for the approximation function

Before we choose the approximation function, we should (roughly) know what properties $P_{\mathrm{r},0}(F)$ and $P_{\mathrm{r},D}(F)$ posses. That sheds light on which kind of function should be adopted as the candidate of fitting function. From Eq. (5.1) (Eq. (5.2)), we know that $P_{\mathrm{r},0}(F)$ ($P_{\mathrm{r},D}(F)$) is monotonically increasing. Put differently, $W_{\mathrm{r},0}(F)$ ($W_{\mathrm{r},D}(F)$), as the derivative of the $P_{\mathrm{r},0}(F)$ ($P_{\mathrm{r},D}(F)$), is non-negative function. In addition, $P_{\mathrm{r},0}(F)$ ($P_{\mathrm{r},D}(F)$) should be a smooth function, i.e., at least the third order derivative of $P_{\mathrm{r},0}(F)$ ($P_{\mathrm{r},D}(F)$) should be continuous. Therefore, we should choose the approximation function which satisfy the properties mentioned above. What is more, the approximation function should posses great flexibility to fit any monotonically increasing function. In the following, we list several candidates which are commonly used as the fitting model, and discuss the pros and cons of each candidate function.

Candidate 1: Polynomial function

A polynomial of degree n is a function of the form

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x^1 + a_0 \ , \tag{5.3}$$

where the a's are real numbers. A polynomial function is called quadratic for $n = 2$, cubic for $n = 3$, and so on. It is often used for approximating other functions. The main advantage is that the approximation using this class of functions can be performed easily. However, a polynomial function is not a reliable approximation for $P_{\mathrm{r},0}(F)$, because the polynomial extrapolation usually yields unusable values, i.e., negative values which are not allowed.

Candidate 2: Fourier series

Fourier series is a way to represent a (wave-like) function as the sum of simple sine waves. The study of Fourier series is a branch of Fourier analysis, which is

widely used in digital analysis. The main benefit of Fourier series is that it is easy to perform computationally. However, analogous to polynomial function, extrapolation using the Fourier series also yields negative values.

Candidate 3: Rational polynomial function

Rational polynomial function, also called rational function, has the form

$$y(x) = \frac{a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x^1 + a_0}{b_m x^m + b_{m-1} x^{m-1} + \cdots + b_2 x^2 + b_1 x^1 + b_0} , \tag{5.4}$$

where n and m are non-negative integers and define the degree of the numerator and denominator, respectively. This class of function is commonly used in the curve fitting, because it can take extremely wide range of shapes and has better interpolation properties than the polynomial functions. However, the unconstrained rational function may not be monotonically increasing, thus the $W_{\mathrm{r},0}(F)$ is negative. In addition, it is usually difficult to find adequate starting values for the fitting.

Candidate 4: Cubic spline interpolation

A spline is numeric function which is piecewise-defined by polynomial functions. A common choice is cubic spline. The flexility of spline interpolation makes sure that the fitting function pass through each point in the list of MC integration values. However, since the size of the sample is finite, the MC integration is not precise enough to distinguish $P_{\mathrm{r},0}(F) \gtrsim 0$ from $P_{\mathrm{r},0}(F) = 0$ or $P_{\mathrm{r},0}(F) \lesssim 1$ from $P_{\mathrm{r},0}(F) = 1$ for some finite ranges. In such case, it is possible that the cubic spline extrapolation yield negative values. See Fig. 5.1 for illustration. Even if we can avoid this negative-value extrapolation by using cubic Hermite interpolating polynomial instead, the derivative of $P_{\mathrm{r},0}(F)$, i.e., $W_{\mathrm{r},0}(F)$ is zero value in a finite range. That should be avoided: Since $W_{\mathrm{r},0}(F)$ is the denominator of $L(D|F)$, the zero values of $W_{\mathrm{r},0}(F)$ result in infinite values of $L(D|F)$. Such a case is not allowed.

Candidate 5: Cumulative distribution function

Since $P_{\mathrm{r},0}(F)$ is monotonically increasing, cumulative distribution function (CDF) is implied as one possible class of fitting functions for approximating $P_{\mathrm{r},0}(F)$. The probability density function, which is the derivative of CDF, is a reasonable choice for approximating $W_{\mathrm{r},0}(F)$ which is positive. Among the CDFs, the regularized incomplete beta function (RIBF) has the great flexibility[1]

$$\begin{aligned} I_{a,b}(x) &= \frac{\mathrm{B}_{a,b}(x)}{\mathrm{B}(a,b)} \\ &= \frac{\int_0^x t^{a-1}\,(1-t)^{b-1}\mathrm{d}t}{\int_0^1 t^{a-1}\,(1-t)^{b-1}\mathrm{d}t}\,, \end{aligned} \tag{5.5}$$

where $\mathrm{B}_{a,b}(x)$ is the incomplete beta function, and $\mathrm{B}(a,b)$ is the beta function

$$\mathrm{B}(a,b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}. \tag{5.6}$$

To increase the flexibility, the approximation function $\widetilde{P_{\mathrm{r},0}}(F)$ consists of several RIBFs

$$\widetilde{P_{\mathrm{r},0}}(F) = w_1\, I_{a_1,b_1}(F) + w_2\, I_{a_2,b_2}(F) + \cdots + w_n\, I_{a_n,b_n}(F)\,, \tag{5.7}$$

with the unit-sum weights $\sum_{i=1}^{n} w_i = 1$. The number of RIBFs depends on the MC integration values. In principle, the more fitting parameters $\{w_i, a_i, b_i\}$ we have, the better is the goodness of the fitting. In practice, however, we use up to six RIBFs for the approximation function $\widetilde{P_{\mathrm{r},0}}(F)$. See Secs. 5.3 and 5.4 for illustration.

[1] Notice that the notation here is analogous to, but different with, those in Eqs. (5.44) and (5.61).

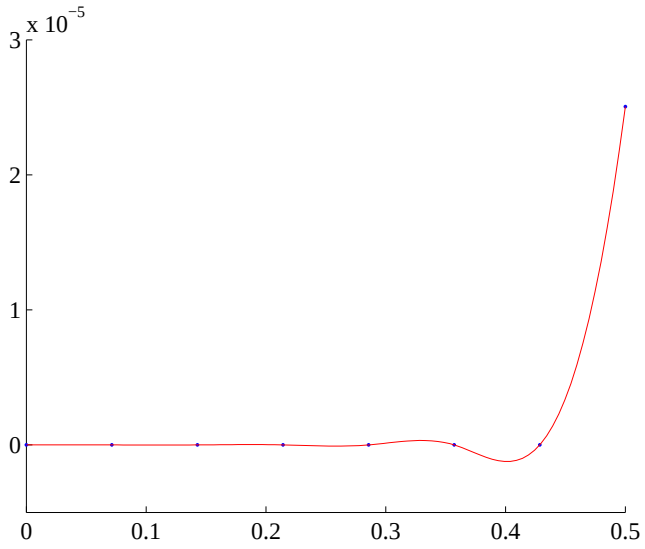

Figure 5.1: The cubic spline extrapolation for an array of values of MC integrations. The blue dots are the array of eight values in which seven values are zero. The red line is the fitting curve of a cubic spline, which yields negative interpolation values.

5.2.3 Iterative algorithm for judging the quality of approximation

After we decide the fitting function for the approximation, here comes one important question: How can we judge the quality of this approximation? Here, we exploit the flexibility offered by Eq. (4.20) for the purpose. Assume we have chosen a certain prior density $W_0(F)$, [2] such that the corresponding prior density of the probability is

$$w_0(p) = \frac{w_\mathrm{r}(p) W_0(f(p))}{\int (\mathrm{d}p')\, w_\mathrm{r}(p')\, \delta\big(f(p) - f(p')\big)}. \tag{5.8}$$

[2] The prior density $W_0(F)$ should be nonzero everywhere, except perhaps at few isolated points.

Therefore, the antiderivative of the integral in Eq. (4.4) yields

$$\begin{aligned} P_0(F) &= \int (\mathrm{d}p)\, w_0(p)\, \eta\big(F - f(p)\big) \\ &= \int (\mathrm{d}p)\, \frac{w_\mathrm{r}(p) W_0\big(f(p)\big)}{W_{\mathrm{r},0}\big(f(p)\big)} \int_0^F \mathrm{d}F'\, \delta\big(F' - f(p)\big) \\ &= \int_0^F \frac{\mathrm{d}F'\, W_0(F')}{W_{\mathrm{r},0}(F')} \int (\mathrm{d}p)\, w_\mathrm{r}(p)\, \delta\big(F' - f(p)\big) = \int_0^F \mathrm{d}F'\, W_0(F') \end{aligned} \tag{5.9}$$

upon recalling Eq. (4.22).

Now we start from the approximation $\widetilde{w}_0(p) = w_\mathrm{r}(p) W_0\big(f(p)\big) \Big/ \widetilde{W}_{\mathrm{r},0}\big(f(p)\big)$, such that the approximation $\widetilde{P}_0(F)$ for $P_0(F)$ is

$$\widetilde{P}_0(F) = \int (\mathrm{d}p)\, \widetilde{w}_0(p)\, \eta\big(F - f(p)\big) = \int_0^F \mathrm{d}F'\, \frac{W_{\mathrm{r},0}(F')}{\widetilde{W}_{\mathrm{r},0}(F')} W_0(F')\,. \tag{5.10}$$

If $\widetilde{P}_0(F) \simeq \int_0^F \mathrm{d}F'\, W_0(F')$ is sufficiently accurate, it means that $\widetilde{W}_{\mathrm{r},0}(F)$ approximates $W_{\mathrm{r},0}(F)$ well. If it is not, an approximation $\widetilde{W}_0(F)$ for $\dfrac{\partial}{\partial F}\widetilde{P}_0(F)$ provides $\widetilde{W}_{\mathrm{r},0}(F)\Big|_\mathrm{new} = \widetilde{W}_{\mathrm{r},0}(F)\widetilde{W}_0(F)/W_0(F)$, which improves on the approximation $\widetilde{W}_{\mathrm{r},0}(F)$. However, we should notice that $\widetilde{W}_{\mathrm{r},0}(F)$ cannot be exactly equal to $W_{\mathrm{r},0}(F)$, because the integral in Eq. (5.10) also requires a MC integration with its intrinsic fluctuations.

In the following, we describe the essence of an iteration algorithm for successive approximations of $W_{\mathrm{r},0}(F)$. Since the F-likelihood $L(D|F)$ does not depend on the prior $W_0(F)$, we can choose $W_0(F) = 1$ so that $P_0(F) = F$ in Eq. (5.9), and the nth iteration of the algorithm consists of these steps:

S1 For given $W_{\mathrm{r},0}^{(n)}(F)$, sample the probability space in accordance with the prior $w_0^{(n)}(p) = w_\mathrm{r}(p)/W_{\mathrm{r},0}^{(n)}\big(f(p)\big)$.

S2 Use this sample for a MC integration of

$$P_0^{(n)}(F) = \int (\mathrm{d}p)\, w_0^{(n)}(p)\, \eta\big(F - f(p)\big)\,.$$

S3 Escape the loop if $P_0^{(n)}(F) \simeq F$ with the desired accuracy.

S4 Fit a suitable several-parameter function to the MC values of $P_0^{(n)}(F)$.

S5 Differentiate this function to obtain $W_0^{(n)}(F) \simeq \dfrac{\partial}{\partial F} P_0^{(n)}(F)$; update $n \to n+1$ and $W_{\mathrm{r},0}^{(n)}(F) \to W_{\mathrm{r},0}^{(n+1)}(F) = W_{\mathrm{r},0}^{(n)}(F) W_0^{(n)}(F)$; return to step S1.

The sampling in step S1 consumes most of the CPU time in each round of iteration. In each round, the step S2 is important, because it updates $W_0^{(n)}(F) \to W_{\mathrm{r},0}^{(n+1)}(F)$ for the next round of iteration. Therefore, for each round, we require the size of the sample to be sufficiently large. Once the accuracy of approximation is close to the one we desire, we can increase the sample size in the last iteration to get a better approximation.

Analogous to the above iterations algorithm for $W_{\mathrm{r},0}(F)$, we compute the numerator $W_{\mathrm{r},D}(F)$ in Eq. (4.21). The same iteration algorithm for works $W_{\mathrm{r},D}(F)$ by simply replacing $W_{\mathrm{r},0}^{(n)}(F) \to W_{\mathrm{r},D}^{(n)}(F)$ and $w_{\mathrm{r}}(p) \to w_{\mathrm{r}}(p) L(D|p)$. Eventually, we get the F-likelihood,

$$L(D|F) = \frac{W_{\mathrm{r},D}(F)}{W_{\mathrm{r},0}(F)}\,, \tag{5.11}$$

and can then proceed to determine the BLIs of Sec. 4.4.

In practice, it is not really necessary to iterate until $P_0^{(n)}(F)$ equals F to a very high accuracy. As long as $W_{\mathrm{r},0}^{(n)}(F)$ is reliable over the whole range from $F = 0$ to $F = 1$, we escape the loop and record the corresponding $w_0^{(n)}(p)$. Then we sample in accordance with the posterior density $\propto w_0^{(n)}(p) L(D|p)$, and fit $W_{\mathrm{r},D}^{(n)}(F)$ to provide a reliable posterior density $W_{\mathrm{r},D}^{(n)}(F)$, and so we obtain the F-likelihood of (5.11). Regarding the fitting of a several-parameter function in step S4, we note that, usually, a truncated Fourier series of the form

$$P_0^{(n)}(F) \simeq F + a_1 \sin(\pi F) + a_2 \sin(2\pi F) + a_3 \sin(3\pi F) + \cdots\,, \tag{5.12}$$

with the amplitudes $a_1, a_2, a_3, \ldots$ as the fitting parameters, is a good choice, possibly modified such that the known properties of $P_0^{(n)}(F)$ are properly taken into account. These matters are illustrated by the examples in Sec. 5.4; see, in particular, Fig. 5.6.

5.3 Example: One qubit

We consider the single-qubit situation as a first illustration. The state of a qubit can be specified as

$$\rho(\boldsymbol{r}) = \frac{1}{2}(\mathbf{1} + \boldsymbol{r} \cdot \boldsymbol{\sigma}), \tag{5.13}$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is the vector of Pauli operators, and $\boldsymbol{r} = (x, y, z)$ is the Bloch vector with $x = \langle\sigma_x\rangle$, $y = \langle\sigma_y\rangle$, and $z = \langle\sigma_z\rangle$. Tetrahedron POM is employed (see Eq. (3.3))

$$\Pi_k = \frac{1}{4}(\mathbf{1} + \boldsymbol{a}_k \cdot \boldsymbol{\sigma}) \quad \text{with } k = 1, 2, 3, 4\,. \tag{5.14}$$

As an example of IC POM, the tetrahedron measurement can uniquely determine the qubit. In other words, there is a unique mapping between probability p and density matrix ρ, namely

$$\boldsymbol{r} = 3\sum_{k=1}^{4} p_k \boldsymbol{a}_k\,. \tag{5.15}$$

This tomographic completeness is useful for our discussion, since it permits both the estimation of a parameter of interest directly from the p_ks, as well as estimating that parameter by first estimating the density operator ρ; see Sec. 5.3.2.

5.3.1 SCIs for fidelity and purity

Here we take two properties of the state for example: the fidelity with respect to some target state, and the normalized purity. Both have values between 0

and 1, so that the concepts and tools of the preceding sections are immediately applicable.

The fidelity[3]

$$\phi = \mathrm{tr}\{|\sqrt{\rho}\,\sqrt{\rho_{\mathrm{tar}}}|\} \tag{5.16}$$

is a measure of overlap between the actual state ρ and the target state ρ_{tar}.

Assume the vector $\boldsymbol{r}$ is the Bloch vector for ρ and $\boldsymbol{t} = \mathrm{tr}\{\boldsymbol{\sigma}\rho_{\mathrm{tar}}\}$ is that for ρ_{tar}, such that the fidelity can be expressed in the following form

$$\phi = \left[\frac{1}{2}(1 + \boldsymbol{r}\cdot\boldsymbol{t}) + \frac{1}{2}\sqrt{1-r^2}\sqrt{1-t^2}\right]^{\frac{1}{2}} \geq \sqrt{\frac{1-t}{2}}\,, \tag{5.17}$$

where $r = |\boldsymbol{r}|$ and $t = |\boldsymbol{t}|$, and the lower bound is reached for $\boldsymbol{r} = -\boldsymbol{t}/t$. If we choose the target state to be $|\mathrm{tar}\rangle = |0\rangle$, then the fidelity is a function of only the z-component of $\boldsymbol{r}$, namely $\phi = \left[\frac{1}{2}(1+z)\right]^{\frac{1}{2}}$.

The purity $\mathrm{tr}\{\rho\}^2$ is a measure of the mixedness of a state, with values between $\frac{1}{2}$ (for the completely mixed state) and 1 (for pure states). To linearly rescale the range from $[\frac{1}{2}, 1]$ to $[0, 1]$, we define the normalized purity by $\gamma = 2\,\mathrm{tr}\{\rho^2\} - 1$, so that $\gamma = r^2$ is simply the squared length of the Bloch vector. By using the relation in Eq. (5.15), we express the fidelity (with respect to $|0\rangle$) and normalized purity in terms of the probabilities p

$$\phi(p) = \sqrt{1-2p_1} \quad \text{and} \quad \gamma(p) = 12\sum_{k=1}^{4} p_k^2 - 3, \tag{5.18}$$

respectively.

The state $\rho = \frac{1}{2}(\mathbf{1} + 0.9\,\sigma_z)$ is used to generate the data in the simulation. This state has fidelity $\Phi = \sqrt{0.95} = 0.9747$ (for target state $|0\rangle$) and normalized purity $\Gamma = 0.81$ — the "true" values for the two parameters to be estimated from

[3] Also called square root fidelity in some literatures.

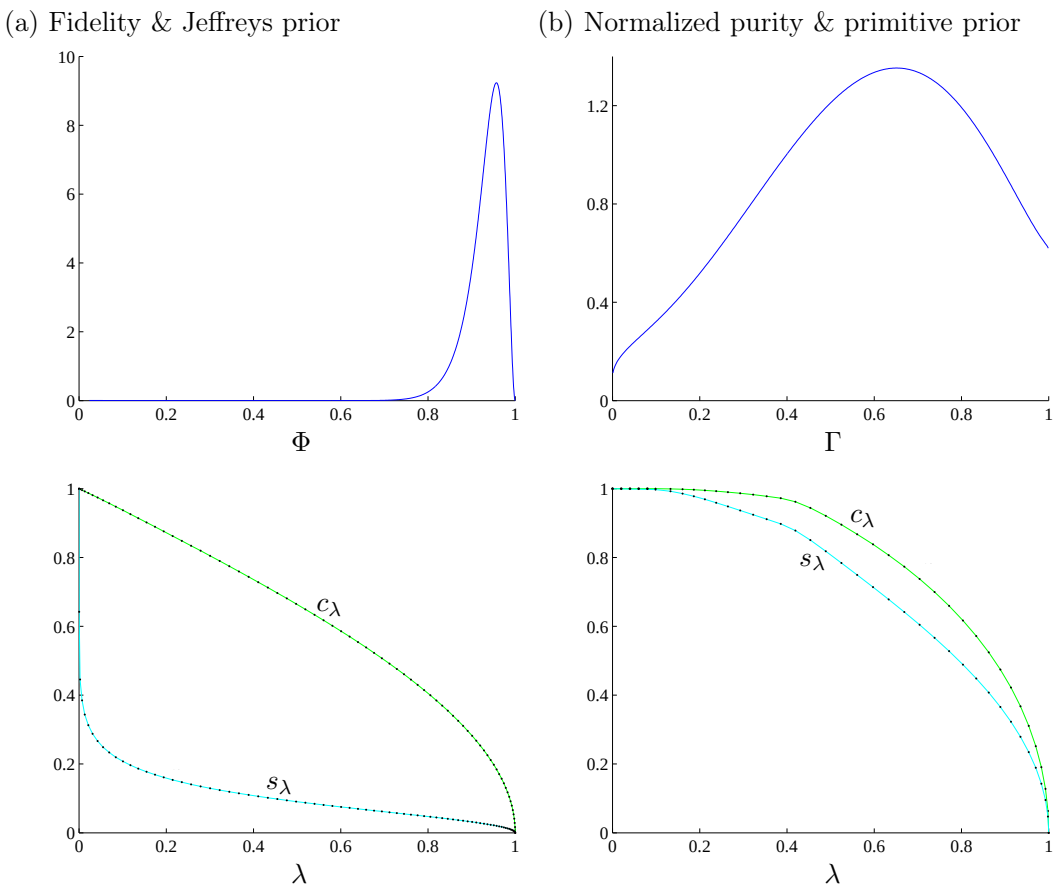

Figure 5.2: Single-qubit fidelity (with respect to $|0\rangle$) and normalized purity for a simulated tetrahedron measurement of 36 copies. Top plots: The Φ-likelihood $L(D|\Phi)$ and the Γ-likelihood $L(D|\Gamma)$ for, respectively, the Jeffreys prior and the primitive prior on the probability space. Bottom plots: The size s_λ (cyan curves) and the credibility c_λ (green curves) for BLIs as functions of λ, for (a) fidelity and the Jeffreys prior; (b) normalized purity and the primitive prior. The black dots mark values obtained from the Hamiltonian Monte Carlo algorithm for evaluating the size and credibility integrals. The cyan lines are fitted to the s_λ values using a Padé approximant, while the green lines for c_λ are obtained from the cyan lines with the aid of Eq. (4.36).

the data[4]. 36 copies of this state are measured by the tetrahedron measurement, and gave data $D = (n_1, n_2, n_3, n_4) = (2, 10, 11, 13)$, where n_k is the number of clicks registered by the detector for measurement outcome Π_k.

[4]Recall that lower-case letters (here ϕ, γ) stand for functions on the probability space, and upper-case letters (here Φ, Γ) are for the function values that label the hypersurfaces.

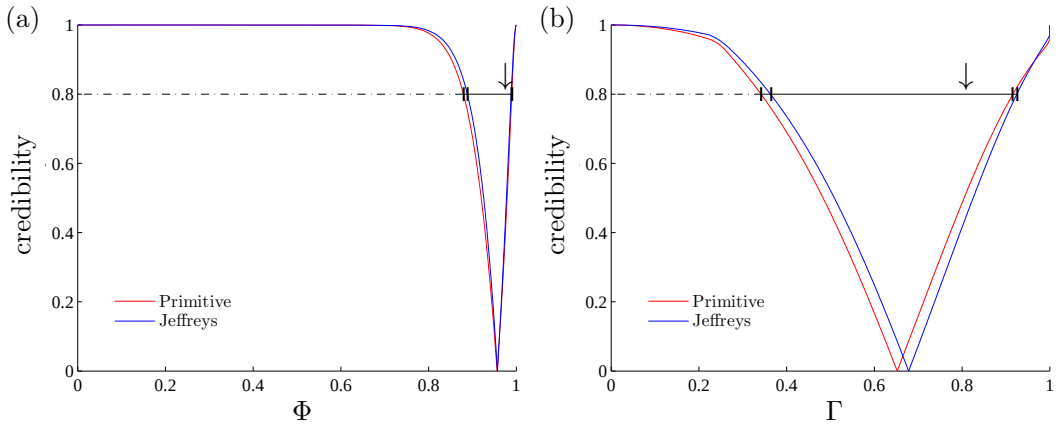

Figure 5.3: Optimal error intervals for (a) fidelity Φ (with respect to $|0\rangle$), and (b) normalized purity Γ, for a qubit state probed with the tetrahedron measurement. The red curves are for the primitive prior; the blue curves are for the Jeffreys prior. These curves delineate the boundaries of the SCIs for different credibility values; the cusps are located at the maximum-likelihood estimates $\widehat{\Phi}_{\text{ML}}$ and $\widehat{\Gamma}_{\text{ML}}$, respectively. For illustration, the SCIs for credibility 0.8 are the intervals indicated by the black bars. The true values of $\Phi = 0.9747$ and $\Gamma = 0.81$, marked by the down-pointing arrows ($\downarrow$), happen to be inside these SCIs; this will be so for 80% of all cases when the simulation is repeated very often for many different true states. Although, only $N = 36$ qubits are measured in the simulated experiment, the SCIs are almost the same for the two priors.

Since the quantum constraint for the single-qubit case is simple, the induced priors $W_0(\Phi)$ and $W_0(\Gamma)$, both for the primitive prior Eq. (3.11) and the Jeffreys prior Eq. (3.12), can be obtained by an analytical evaluation of Eq. (4.22). The top plots in Fig. 5.2 report the F-likelihoods $L(D|\Phi)$ and $L(D|\Gamma)$ thus obtained for the Jeffreys prior and the primitive prior, respectively.

The size s_λ and the credibility c_λ for the resulting BLIs, computed from these F-likelihoods together with the respective induced priors, are shown in the bottom plots in Fig. 5.2. The dots mark values obtained by numerical integration that employs the Hamiltonian Monte Carlo algorithm [102]. The green curves through the credibility points, which is obtained by integrating over the cyan curve fitted to the size points, demonstrate the relation in Eq. (4.36) between s_λ and c_λ.

In Fig. 5.3, we report the SCIs for fidelity Φ and normalized purity Γ, both for the primitive prior (red lines) and for the Jeffreys prior (blue lines). The horizontal interval between the two branches of the curves is the SCI with a specific credibility; see the examples for $c = 0.8$ marked on the plots. In Fig. 5.3, we observe that even for $N = 36$ copies of the state measured, the choice of prior has little effect on the SCIs. Put differently, our conclusions are dominated by the data (not necessarily many), not by the choice of prior.

5.3.2 Direct and indirect estimation of state properties

In Sec. 4.1, we mentioned that the best guess for the propertiies of interest may not, and often does not, come from the best guess for the quantum state (see [73]). In the context of error intervals, we compare the two approaches for our qubit example. On the one hand, we construct the error intervals by directly estimating the value of the property from the data, as we have done in the previous section. On the other hand, we first construct the error regions (SCRs specifically; see [1]) for the quantum state, then the error interval for the desired property is given by the range of property values for the states contained in the error region of states (see Fig. 4.1 for illustration). We refer to the two respective approaches as direct and indirect state-property estimation, with the abbreviations of DSPE and ISPE.

In Fig. 5.4, the ORIs for fidelity Φ and normalized purity Γ for the single-qubit data of Figs. 5.2 and 5.3 are demonstrated. Red curves are obtained via DSPE, while purple curves are via ISPE. Here, we employ the primitive prior Eq. (3.11) as the prior density $w_0(p)$ on the probability space, together with the induced prior densities $W_0(\Phi)$ and $W_0(\Gamma)$ for the fidelity and the normalized purity. Obviously, the error intervals obtained by these two approaches are quite different in this situation. In particular, as we observe from Fig. 5.4, DSPE

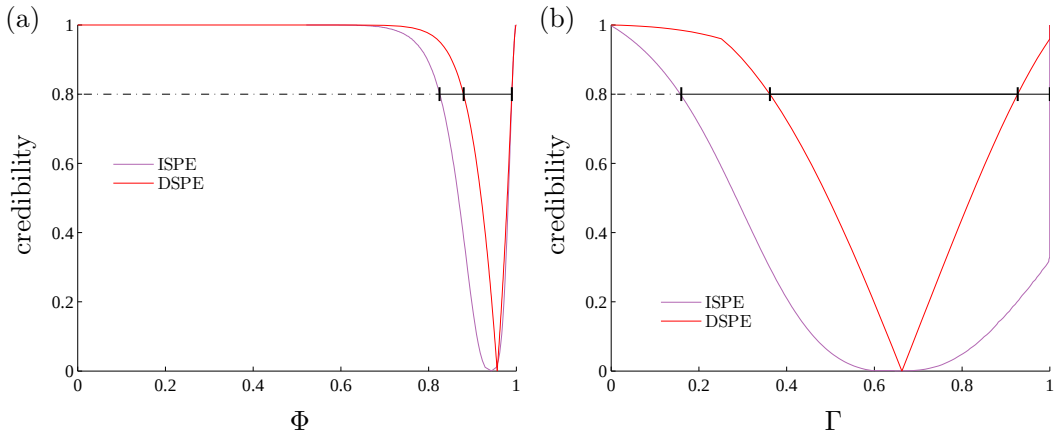

Figure 5.4: Direct and indirect state-property estimation: Error intervals for (a) fidelity Φ and (b) normalized purity Γ by DSPE (red curves) and ISPE (purple curves), for the same simulated data as in Figs. 5.2 and 5.3. As in Fig. 5.3, the horizontal bars indicate the intervals for credibility 0.8. Consistent with what the sketch in Fig. 4.1 suggests, the intervals obtained from ISPE are larger than the actual SCIs that result from proper DSPE. In plot (a), one can also clearly see that the maximum-likelihood fidelity $\widehat{\Phi}_{\mathrm{ML}}$ is not the fidelity of the maximum-likelihood state $\widehat{\rho}_{\mathrm{ML}}$: The cusps of the red and purple curves are at different Φ values.

reports smaller intervals than ISPE does. Note that, however, the meaning of two intervals obtained via ISPE and DSPE are quite different. For DSPE, the credibility value of the corresponding ORI is the posterior content of that interval for the property itself. For ISPE, however, the credibility is the posterior content for the *state* error region. In general, there is no simple relation to the probability of containing the true property value. Figure 4.1 illustrates the above conclusion, where the range of F values across the SCR is larger than the range of the SCI.

5.4 Example: Two qubits

5.4.1 CHSH quantity, TAT scheme, and simulated experiment

In the two-qubit states case, we are usually interested in the entangled states. As an important quantity to judge whether a quantum state is entangled or not, Clauser-Horne-Shimony-Holt (CHSH) quantity [103,104] is chosen to be the property to be estimated in this section. Mathematically, the CHSH quantity is specified in the following

$$\theta = \mathrm{tr}\{(A_1 \otimes B_1 + A_2 \otimes B_1 + A_1 \otimes B_2 - A_2 \otimes B_2)\rho\}\,, \tag{5.19}$$

where $A_j = \boldsymbol{a}_j \cdot \boldsymbol{\sigma}$ and $B_{j'} = \boldsymbol{b}_{j'} \cdot \boldsymbol{\sigma}$ with unit vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{b}_1$, and $\boldsymbol{b}_2$ are components of the Pauli vector operators for the two qubits. The maximum value of $|\theta|$ is $\sqrt{8}$. If $|\theta| > 2$, the two-qubit state is surely entangled. Therefore, it is important to distinguish reliably between $|\Theta| < 2$ and $|\Theta| > 2$. [5]

A common choice for the single-qubit observables is

$$A_1 = \sigma_x\,, \quad A_2 = \sigma_z\,, \quad \left.\begin{matrix} B_1 \\ B_2 \end{matrix}\right\} = -\frac{1}{\sqrt{2}}(\sigma_x \pm \sigma_z)\,, \tag{5.20}$$

such that

$$\theta = -\sqrt{2}\langle \sigma_x \otimes \sigma_x + \sigma_z \otimes \sigma_z \rangle\,. \tag{5.21}$$

The limiting values $\theta = \pm\sqrt{8}$ are obtained by choosing two of the "Bell states", *viz.* the maximally entangled states $\rho = \frac{1}{4}(\mathbf{1} \mp \sigma_x \otimes \sigma_x)(\mathbf{1} \mp \sigma_z \otimes \sigma_z)$, the common eigenstates of $\sigma_x \otimes \sigma_x$ and $\sigma_z \otimes \sigma_z$ with same eigenvalue -1 or $+1$.

[5] Recall footnote 4: The distinction between θ (lower case) and Θ (upper case) is analogous to that between ϕ, γ and Φ, Γ.

Since the CHSH quantity Θ only involves the xz planes of the two Bloch balls, but no y components, we employ the trine-antitrine (TAT) POM (see [100] and Sec. 6 in [1]) to determine the CHSH quantity. Qubit 1 is measured by the trine-outcome POM with outcome operators

$$\Pi_1^{(1)} = \frac{1}{3}\left(1+\sigma_z\right), \qquad \left.\begin{matrix}\Pi_2^{(1)}\\ \Pi_3^{(1)}\end{matrix}\right\} = \frac{1}{3}\left(1 \pm \frac{\sqrt{3}}{2}\sigma_x - \frac{1}{2}\sigma_z\right), \tag{5.22}$$

and the $\Pi_{j'}^{(2)}$s for qubit 2 have the signs of σ_x and σ_z reversed. The nine probability operators of the product POM are

$$\Pi_k = \Pi_j^{(1)} \otimes \Pi_{j'}^{(2)} \quad \text{with} \quad k = 3(j-1)+j' \equiv [jj'], \tag{5.23}$$

that is $1 = [11]$, $2 = [12]$, $\ldots$, $5 = [22]$, $\ldots$, $8 = [32]$, $9 = [33]$, and Eq. (5.21) is expressed in terms of the TAT probabilities p

$$\theta(p) = \sqrt{8}\Big[3(p_1+p_5+p_9)-1\Big], \tag{5.24}$$

where the TAT probabilities are given by the Born rule $p_i = \mathrm{tr}\{\Pi_i\rho\}$.

For the CHSH quantity in Eq. (5.24) which is a linear function of ps, we choose the vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ to be fixed such that the Pauli vector operators A_1, A_2, B_1, B_2 has the form in Eq. (5.24). In fact, we can orient the vectors such that θ is largest for the given ρ

$$\theta_{\text{opt}} = 2\Big[\langle\sigma_x\otimes\sigma_x\rangle^2 + \langle\sigma_x\otimes\sigma_z\rangle^2 + \langle\sigma_z\otimes\sigma_x\rangle^2 + \langle\sigma_z\otimes\sigma_z\rangle^2\Big]^{\frac{1}{2}}, \tag{5.25}$$

here, we denote by θ_{opt} the optimized CHSH quantity. In terms of the TAT

probabilities, it is given by

$$\left(\frac{\theta_{\rm opt}(p)}{4}\right)^2 = 9\sum_{k=1}^{9} p_k^2 - 3\Big[(p_1+p_2+p_3)^2+(p_4+p_5+p_6)^2+(p_7+p_8+p_9)^2 + (p_1+p_4+p_7)^2+(p_2+p_5+p_8)^2+(p_3+p_6+p_9)^2\Big] + 1 \quad (5.26)$$

The optimal-vectors CHSH quantity is a quadratic function of TAT probabilities, though the fixed-vectors CHSH quantity in Eq. (5.24) is a linear function of ps.

Clearly, the inequality $|\theta| \leq \theta_{\rm opt}$ holds for any two-qubit state ρ. As an extreme example, the Bell states of the form $\rho = \frac{1}{4}(\mathbf{1} \pm \sigma_x \otimes \sigma_x)(\mathbf{1} \mp \sigma_z \otimes \sigma_z)$, which are the common eigenstates of $\sigma_x \otimes \sigma_x$ and $\sigma_z \otimes \sigma_z$ with opposite eigenvalues, give us $\theta = 0$ and $\theta_{\rm opt} = \sqrt{8}$. The same values are also found for other states, among them all four common eigenstates of $\sigma_x \otimes \sigma_z$ and $\sigma_z \otimes \sigma_x$.

In the simulation, we use the following true state

$$\rho_{\rm true} = \frac{1}{4}(\mathbf{1} - x\sigma_x \otimes \sigma_x - y\sigma_y \otimes \sigma_y - z\sigma_z \otimes \sigma_z) \quad (5.27)$$

with $(x, y, z) = \frac{1}{20}(18, -15, -14)$, for which the TAT probabilities are

$$\begin{pmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \\ p_7 & p_8 & p_9 \end{pmatrix} = \frac{1}{60}\begin{pmatrix} 2 & 9 & 9 \\ 9 & 10 & 1 \\ 9 & 1 & 10 \end{pmatrix} \quad (5.28)$$

and the true values of Θ and $\Theta_{\rm opt}$ are

$$\Theta = \sqrt{2}(x+z) = \frac{1}{5}\sqrt{2} = 0.2828\,,$$
$$\Theta_{\rm opt} = 2\sqrt{x^2+z^2} = \sqrt{\frac{26}{5}} = 2.2804\,. \quad (5.29)$$

$N = 180$ copies of the state are measured, with the detector clicks $D =$

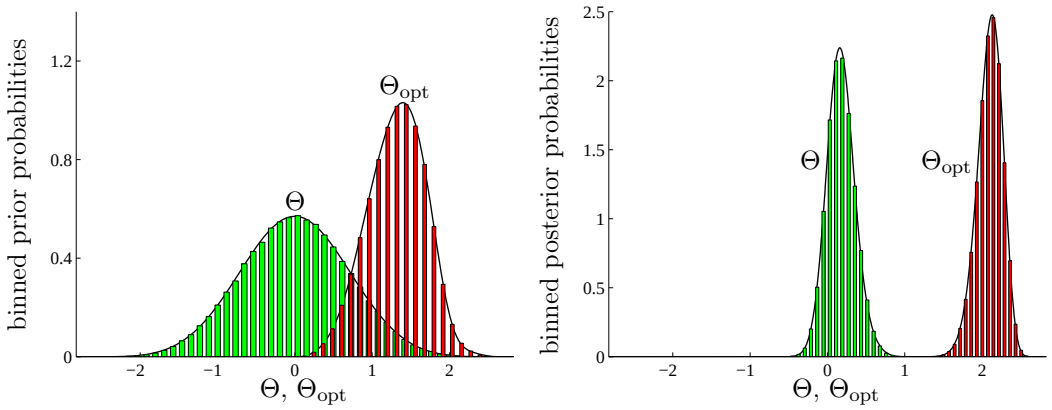

Figure 5.5: Left: Histogram of CHSH values in a random sample of 500 000 states in accordance with the primitive prior of Eq. (3.11). For Θ of Eq. (5.24) we have the full range of $-\sqrt{8} \leq \Theta \leq \sqrt{8}$, whereas Θ_{opt} of Eq. (5.26) is positive by construction. — Right: Corresponding histogram for a random sample drawn from the posterior distribution for the simulated data in Eq. (5.30). — On the left, the black-line envelopes show the few-parameter approximations of Eq. (5.31) with Eq. (5.32) for Θ and Eq. (5.62)–Eq. (5.63) for Θ_{opt}. On the right, the envelopes are the derivatives of the fits to $P_{\text{r},D}(\Theta)$ and $P_{\text{r},D}(\Theta_{\text{opt}})$.

$(n_1, n_2, \cdots, n_9) = (9, 28, 30, 28, 27, 3, 29, 1, 25)$. Then the relative frequencies are

$$\frac{1}{180}\begin{pmatrix} 9 & 28 & 30 \\ 28 & 27 & 3 \\ 29 & 1 & 25 \end{pmatrix} = \begin{pmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \\ p_7 & p_8 & p_9 \end{pmatrix} + \frac{1}{180}\begin{pmatrix} 3 & 1 & 3 \\ 1 & -3 & 0 \\ 2 & -2 & -5 \end{pmatrix}, \tag{5.30}$$

Substituting the TAT probabilities by the relative frequencies in Eq. (5.24) and Eq. (5.26), we obtain the estimates for Θ and Θ_{opt} are $\sqrt{2}/30 = 0.0471$ and $16\sqrt{39}/45 = 2.2204$, respectively.

The linear inversion is advocated in [73] as one method to supplement the estimates with error bars that refer to confidence intervals, but the approach has well-known problems [105]. Instead, SCIs for Θ and Θ_{opt} are reported in this thesis, and for those we need the Θ-likelihoods $L(D|\Theta)$ and $L(D|\Theta_{\text{opt}})$. For the sake of convenience, we use Θ and Θ_{opt} themselves, rather than $F = \frac{1}{2}\big(\Theta/\sqrt{8} + 1\big)$ or $F = \Theta_{\text{opt}}/\sqrt{8}$ which have values in the range $0 \leq F \leq 1$, as the properties to

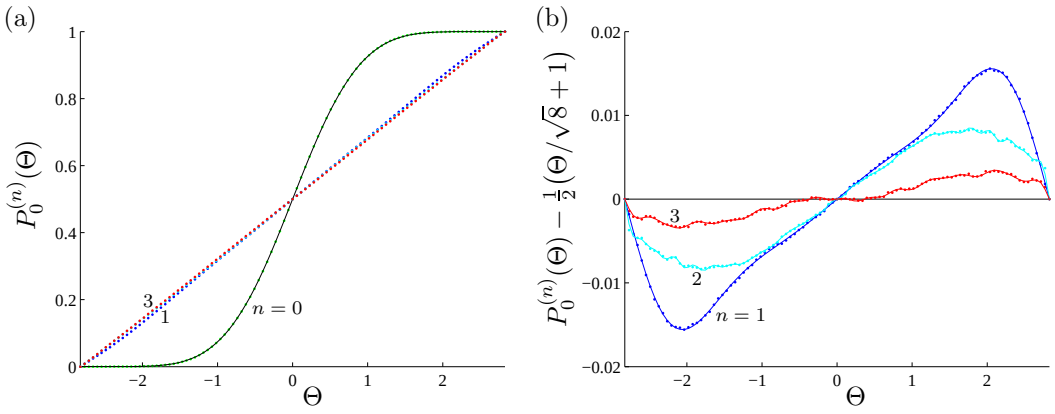

Figure 5.6: Consecutive functions $P_0^{(n)}(\Theta)$ for $n = 0, 1, 2, 3$ as obtained by MC integration. The green dots ($n = 0$) represent values for $P_0(\Theta)$, computed with the primitive prior. The flat regions near the end points at $\Theta = \pm\sqrt{8}$ are a consequence of the $\frac{11}{2}$ power in Eq. (5.44). The black curve through the green dots is the graph of the four-parameter approximation $P_0^{(0)}(\Theta)$ of Eq. (5.31). The blue, cyan, and red dots are the MC values for $n = 1$, 2, and 3, respectively, all close to the straight line $\Theta \mapsto \frac{1}{2}\big(\Theta/\sqrt{8}+1\big)$. The cyan dots are difficult to see between the blue and red dots in plot (a). They are well visible in plot (b), where the straight-line values are subtracted. The curves through the dots in plot (b) show the few-term Fourier approximations analogous to Eq. (5.12).

be estimated. For the MC integration of $P_0(\Theta)$, say, we sample the probability space with the Hamiltonian MC algorithm described in Sec. 4.3 in [102][6].

Sampling the probability space in accordance with the primitive prior of Eq. (3.11), the the distribution of Θ and Θ_{opt} values are shown by the histograms on the left in Fig. 5.5. On the right, the histograms are depicted for a corresponding sample drawn from the posterior distribution with data of Eq. (5.30). It is obvious that the prior distributions contain few values with $\Theta_{\text{opt}} > 2$ and much fewer with $|\Theta| > 2$. For the posterior distributions, values exceeding 2 are prominent for Θ_{opt}, but virtually non-existent for Θ.

Different from the examples of the properties of the single-qubit states, in

[6]The sample probabilities carry a weight proportional to the range of permissible values for $\langle(\sigma_x \otimes \sigma_x)(\sigma_z \otimes \sigma_z)\rangle = -\langle\sigma_y \otimes \sigma_y\rangle$, i.e., parameter q in Eq. (5.48). It is expedient to generate an unweighted sample by resampling ("bootstrapping") the weighted sample. The unweighted sample is then used for the MC integration.

which the $W_{\mathrm{r},0}(F)$ and $W_{\mathrm{r},D}(F)$ can be integrated analytically, the integrals of $W_{\mathrm{r},0}(\Theta)$ ($W_{\mathrm{r},0}(\Theta_{\mathrm{opt}})$) and $W_{\mathrm{r},D}(\Theta)$ ($W_{\mathrm{r},D}(\Theta_{\mathrm{opt}})$) can only be computed numerically. More importantly, to report a proper Θ-likelihood, the power laws near the boundaries is required. In the next two subsections, we will discuss the power laws of the intervals near boundaries for Θ and Θ_{opt}, respectively.

5.4.2 Prior-content function $P_0(\Theta)$ near $\Theta = \pm\sqrt{8}$

In this subsection, we consider the sizes of the regions with $\theta(p) \gtrsim -\sqrt{8}$ and $\theta(p) \lesssim \sqrt{8}$. Take into account the symmetry property $W_0(\Theta) = W_0(-\Theta)$ or $P_0(\Theta) + P_0(-\Theta) = 1$, we use the four-parameter approximation, which is analogous to the approximation for the properties of range $0 \le F \le 1$ (see Eq. (5.7)), as the fitting function for Θ

$$P_0(\Theta) \simeq P_0^{(0)}(\Theta) = w_1 B_{\alpha_1}(\Theta) + w_2 B_{\alpha_2}(\Theta) + w_3 B_{\alpha_3}(\Theta)\,, \tag{5.31}$$

where

$$B_\alpha(\Theta) = \left(\frac{1}{32}\right)^{\alpha+\frac{1}{2}} \frac{(2\alpha+1)!}{(\alpha!)^2} \int_{-\sqrt{8}}^{\Theta} \mathrm{d}x \left(8 - x^2\right)^\alpha \tag{5.32}$$

is a normalized incomplete beta function integral with $B_\alpha(-\sqrt{8}) = 0$ and $B_\alpha(\sqrt{8}) = 1$, and α_1 is the dominating power near the boundaries. The fitting parameters α_2 and α_3 are larger than α_1, while w_1, w_2, w_3 are weights with unit sum .

Now we identify the power α_1 in the following. We denote the kets of the maximally entangled states with $\theta = \pm\sqrt{8}$ by $|\pm\rangle$, that is

$$|+\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}} \quad \text{and} \quad |-\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}\,, \tag{5.33}$$

where $|01\rangle = |0\rangle \otimes |1\rangle$, for example, has $\sigma_z = 1$ for the first qubit and $\sigma_z = -1$

for the second. Since

$$\sigma_x \otimes \mathbf{1}|\pm\rangle = \mp\mathbf{1} \otimes \sigma_x|\pm\rangle \quad \text{and} \quad \sigma_z \otimes \mathbf{1}|\pm\rangle = \mp\mathbf{1} \otimes \sigma_z|\pm\rangle\,, \tag{5.34}$$

we have [recall Eq. (5.23)]

$$\begin{aligned}
\Pi_k|\pm\rangle &= \Pi_{[jj']}|\pm\rangle = \Pi_j^{(1)} \otimes \Pi_{j'}^{(2)}|\pm\rangle \\
&= \frac{1}{9}(1 + \boldsymbol{t}_j \cdot \boldsymbol{\sigma}) \otimes (1 - \boldsymbol{t}_{j'} \cdot \boldsymbol{\sigma})|\pm\rangle \\
&= \frac{1}{9}(1 + \boldsymbol{t}_j \cdot \boldsymbol{\sigma})(1 \pm \boldsymbol{t}_{j'} \cdot \boldsymbol{\sigma}) \otimes \mathbf{1}|\pm\rangle \\
&= \frac{1}{9}\Big[(1 \pm \boldsymbol{t}_j \cdot \boldsymbol{t}_{j'})\mathbf{1} + (\boldsymbol{t}_j \pm \boldsymbol{t}_{j'} \pm \mathrm{i}\boldsymbol{t}_j \times \boldsymbol{t}_{j'}) \cdot \boldsymbol{\sigma}\Big] \otimes \mathbf{1}|\pm\rangle\,,
\end{aligned} \tag{5.35}$$

where

$$\boldsymbol{t}_1 = \boldsymbol{e}_z\,, \quad \boldsymbol{t}_2 = \frac{\sqrt{3}}{2}\boldsymbol{e}_x - \frac{1}{2}\boldsymbol{e}_z\,, \quad \boldsymbol{t}_3 = -\frac{\sqrt{3}}{2}\boldsymbol{e}_x - \frac{1}{2}\boldsymbol{e}_z \tag{5.36}$$

are the three unit vectors of the trine.

Now let Θ deviate from $\pm\sqrt{8}$, the corresponding changes for the state has the form

$$\begin{aligned}
\rho_\epsilon &= \frac{\big(|\pm\rangle\langle\pm| + \epsilon A^\dagger\big)\big(|\pm\rangle\langle\pm| + \epsilon A\big)}{1 + \epsilon\langle\pm|(A^\dagger + A)|\pm\rangle + \epsilon^2 \mathrm{tr}\{A^\dagger A\}} \\
&= |\pm\rangle\langle\pm| + \epsilon\,(A_\pm^\dagger + A_\pm) + O(\epsilon^2)\,,
\end{aligned} \tag{5.37}$$

where ϵ is infinitesimal and A an arbitrary two-qubit operator. A particular traceless rank-1 operator

$$A_\pm = |\pm\rangle\langle\pm|A\big(\mathbf{1} - |\pm\rangle\langle\pm|\big) \tag{5.38}$$

is central to the following derivation because of its properties $A_\pm|\pm\rangle = 0$ and

$|\pm\rangle\langle\pm|A_\pm = A_\pm$. Then, the TAT probabilities for ρ_ϵ are

$$\begin{aligned} p_k &= \mathrm{tr}\{\rho_\epsilon \Pi_k\} = \langle\pm|\Pi_k|\pm\rangle + \epsilon\,\mathrm{tr}\Big\{(A_\pm^\dagger + A_\pm)\Pi_k\Big\} + O(\epsilon^2) \\ &= \frac{1}{9}\big(1 \pm \boldsymbol{t}_j \cdot \boldsymbol{t}_{j'}\big) + \epsilon\Big[(\boldsymbol{t}_j \pm \boldsymbol{t}_{j'}) \cdot \boldsymbol{\alpha}_\pm \mp (\boldsymbol{t}_j \times \boldsymbol{t}_{j'}) \cdot \boldsymbol{\beta}_\pm\Big] + O(\epsilon^2) \quad (5.39) \end{aligned}$$

with the real vectors $\boldsymbol{\alpha}_\pm$ and $\boldsymbol{\beta}_\pm$ given by

$$\frac{2}{9}\mathrm{tr}\big\{\boldsymbol{\sigma} \otimes \mathbf{1} A_\pm\big\} = \boldsymbol{\alpha}_\pm + \mathrm{i}\boldsymbol{\beta}_\pm\,. \quad (5.40)$$

Due to the trine geometry, the x and z components of $\boldsymbol{\alpha}_\pm$ and the y component of $\boldsymbol{\beta}_\pm$ matter, while the other three components do not. In the eight-dimensional probability space, then, we have increments $\propto \epsilon$ in three directions only, and increments $\propto \epsilon^2$ in the other five directions. For the primitive prior, therefore, the size of the ϵ-vicinity is $\propto \epsilon^{3\times1+5\times2} = \epsilon^{13}$.

The sum of probabilities in Eq. (5.24) is

$$p_1 + p_5 + p_9 = p_{[11]} + p_{[22]} + p_{[33]} = \frac{1}{3}(1 \pm 1) + O(\epsilon^2)\,, \quad (5.41)$$

so that $\Theta = \pm\sqrt{8}\,[1 - O(\epsilon^2)]$ or $\sqrt{8} - |\Theta| \propto \epsilon^2$. Therefore, the prior content $P_0(\Theta)$ near the boundaries

$$P_0(\Theta) \propto \Big(\sqrt{8} + \Theta\Big)^{\frac{13}{2}} \quad \text{near } \Theta = -\sqrt{8} \quad (5.42)$$

and

$$1 - P_0(\Theta) \propto \Big(\sqrt{8} - \Theta\Big)^{\frac{13}{2}} \quad \text{near } \Theta = \sqrt{8}\,. \quad (5.43)$$

As the derivative of $P_0(\Theta)$, $W_0(\Theta)$ has the denominating power $\alpha_1 = \frac{11}{2}$ near $\pm\sqrt{8}$

$$\frac{\mathrm{d}}{\mathrm{d}\Theta}P_0(\Theta) = W_0(\Theta) \propto \Big(\sqrt{8} - |\Theta|\Big)^{\frac{11}{2}} \quad \text{for} \quad |\Theta| \lesssim \sqrt{8}. \quad (5.44)$$

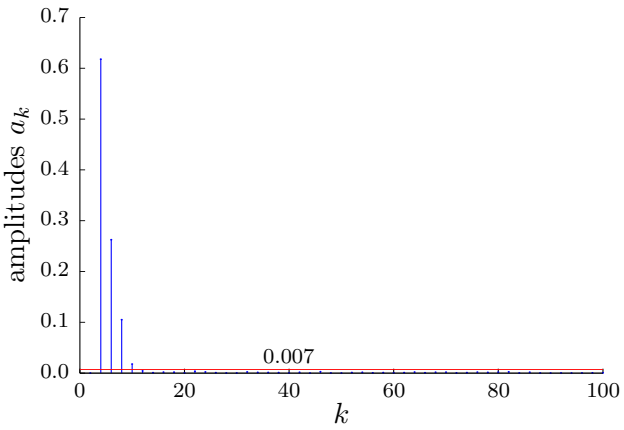

Figure 5.7: Fourier coefficients of Eq. (5.12) for $P_0^{(1)}(\Theta)$ ($\equiv$ blue dots in Fig. 5.6). The amplitudes with values below the threshold value of 0.007 (red line) are set to zero by the "low-pass filter" in order to remove the high-frequency noise in $P_0^{(1)}(\Theta)$. All amplitudes with odd index vanish, $a_1 = a_3 = a_5 = \cdots = 0$, and only the four amplitudes a_4, a_6, a_8, and a_{10} remain nonzero.

The green dots in Fig. 5.6(a) represent the $P_0(\Theta)$ values obtained with the sample of 500 000 sets of permissible probabilities which generated the histograms in Fig. 5.5(left). The black curve through the green dots in 5.6(a) is $P_0^{(0)}(\Theta)$ with the fitting parameters $\alpha_2 = \alpha_1 + 1.6700$, $\alpha_3 = \alpha_1 + 5.4886$, and $(w_1, w_2, w_3) = (0.4691, 0.2190, 0.3119)$. The black envelope for the green Θ histogram in 5.5(left) is the corresponding approximation for $W_0(\Theta)$.

5.4.3 Iterated MC integrations for $P_0(\Theta)$

Employing the iteration algorithm in Sec. 5.2.3, and taking into account into account the power law near the boundaries, we obtain the subsequent approximations $P_0^{(1)}(\Theta)$, $P_0^{(2)}(\Theta)$, and $P_0^{(3)}(\Theta)$, which are shown as the blue, cyan, and red dots in Fig. 5.6(a), and after subtracting $\frac{1}{2}\big(\Theta/\sqrt{8}+1\big)$, also in 5.6(b). The truncated Fourier series of Eq. (5.12) with $F = \frac{1}{2}(\Theta/\sqrt{8}+1)$ are used for fitting a smooth curve to the noisy MC values for $P_0^{(1)}(\Theta)$, $P_0^{(2)}(\Theta)$, and $P_0^{(3)}(\Theta)$. As a consequence of the symmetry of $P_0(\Theta) + P_0(-\Theta) = 1$, all Fourier amplitudes a_k with odd k vanish.

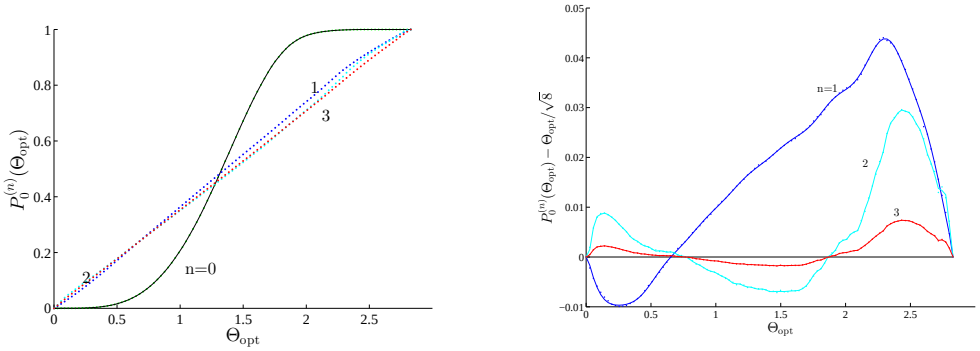

Figure 5.8: Consecutive functions $P_0^{(n)}(\Theta_{\rm opt})$ for $n = 0, 1, 2, 3$ as obtained by MC integration. Analogous to Fig. 5.6, the green dots ($n = 0$) represent values for $P_0(\Theta_{\rm opt})$, computed with the primitive prior, while the blue, cyan, and red dots are the MC values for $n = 1$, 2, and 3, respectively. which are all close to the straight line $\Theta_{\rm opt} \mapsto \Theta_{\rm opt}/\sqrt{8}$. The flat regions near the end points at $\Theta_{\rm opt} = 0$ and $\Theta_{\rm opt} = \sqrt{8}$ are a consequence of the 4 and 6 power in Eq. (5.53) and Eq. (5.60), respectively. The black curve through the green dots is the graph of the five-term approximation $P_0^{(0)}(\Theta_{\rm opt})$ of Eq. (5.62). After the subtraction, blue, cyan and red dots are well visible in the right plot. The curves through the dots in the right plot show the few-term Fourier approximations analogous to Eq. (5.12).

For an illustration of the method, we report in Fig. 5.7 the amplitudes a_k of a full Fourier interpolation between the blue dots ($n = 1$) in Fig. 5.6(b). A low-pass filter removes all components with amplitudes below the threshold value set at 0.007 and retains only four nonzero amplitudes. The resulting truncated Fourier series gives the smooth blue curve through the blue dots. Its derivative contributes a factor $W_0^{(1)}(F)$ to the reference prior density $W_{\rm r,0}(F)$, in accordance with step S5 of the iteration algorithm in Sec. 5.2. In the next round we treat $P_0^{(2)}(\Theta)$ in the same way, followed by $P_0^{(3)}(\Theta)$ in the third round.

5.4.4 Prior-content function $P_0(\Theta_{\rm opt})$ near $\Theta_{\rm opt} = 0, \sqrt{8}$

In this subsection, we consider the sizes of the intervals with $\Theta_{\rm opt} \gtrsim 0$ and $\Theta_{\rm opt} \lesssim \sqrt{8}$. We wish to establish the $\Theta_{\rm opt}$ analogs of Eq. (5.44) and Eq. (5.31). However, since there is no symmetry property for $\Theta_{\rm opt}$, i.e., $W(\Theta_{\rm opt}) \neq W(\sqrt{8} -$

Θ_{opt}), the powers near boundaries $\Theta_{\text{opt}} \lesssim \sqrt{8}$ and $\Theta_{\text{opt}} \lesssim \sqrt{8}$ are no longer equal to each other.

From Eqs. (5.22) and (5.22) we know that the TAT probabilities $p_1, p_2, \ldots, p_9$ are linear combinations of the expectation values of the single-qubit and two-qubit observables $\tau_i \otimes \tau_j$, where $\boldsymbol{\tau} = (\mathbf{1}, \sigma_x, \sigma_z)$. Switching from $p_1, p_2, \ldots, p_9$ to these expectation values of the eight single-qubit and two-qubit observables, we have

$$\begin{pmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \\ p_7 & p_8 & p_9 \end{pmatrix} \overset{\text{linear}}{\underset{\text{relation}}{\longleftrightarrow}} \left[\begin{array}{c|cc} & \langle \mathbf{1} \otimes \sigma_x \rangle & \langle \mathbf{1} \otimes \sigma_z \rangle \\ \hline \langle \sigma_x \otimes \mathbf{1} \rangle & \langle \sigma_x \otimes \sigma_x \rangle & \langle \sigma_x \otimes \sigma_z \rangle \\ \langle \sigma_z \otimes \mathbf{1} \rangle & \langle \sigma_z \otimes \sigma_x \rangle & \langle \sigma_z \otimes \sigma_z \rangle \end{array}\right] \equiv \left[\begin{array}{c|cc} & x_3 & x_4 \\ \hline x_1 & y_1 & y_2 \\ x_2 & y_3 & y_4 \end{array}\right]. \tag{5.45}$$

The corresponding Jacobian matrix is a constant number, such that the integrals in the probability spaces can be transferred to that in the xy space, so is the constraint

$$(\mathrm{d}\rho) = (\mathrm{d}p) = (\mathrm{d}x)\,(\mathrm{d}y)\, w_{\text{cstr}}(x, y)\,. \tag{5.46}$$

For the primitive prior, we have

$$(\mathrm{d}x) = \mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3\,\mathrm{d}x_4 \quad \text{and} \quad (\mathrm{d}y) = \mathrm{d}y_1\,\mathrm{d}y_2\,\mathrm{d}y_3\,\mathrm{d}y_4\,, \tag{5.47}$$

and $w_{\text{cstr}}(x, y)$ equals a normalization factor for permissible values of $x = (x_1, x_2, x_3, x_4)$ and $y = (y_1, y_2, y_3, y_4)$, whereas $w_{\text{cstr}}(x, y) = 0$ for unphysical values. The quantum constraint results from the positivity of the density matrix $\rho \geq 0$. Rewritten the density matrix in terms of x, y and q, of which the range is $-1 \leq q \leq 1$. Then the constraint is satisfied with the permissible values of x

and y determined by [7]

$$\begin{pmatrix} 1+x_1+x_3+y_1 & x_2+y_3 & x_4+y_2 & y_4-q \\ x_2+y_3 & 1-x_1+x_3-y_1 & y_4+q & x_4-y_2 \\ x_4+y_2 & y_4+q & 1+x_1-x_3-y_1 & x_2-y_3 \\ y_4-q & x_4-y_2 & x_2-y_3 & 1-x_1-x_3+y_1 \end{pmatrix} \geq 0\,. \tag{5.48}$$

Of course, it is difficult to express Eq. (5.48) in an explicit way. However, in the following, we can determine the constraint for special cases.

5.4.4.1 The vicinity of $\Theta_{\text{opt}} = 0$

Equation (5.25) implies that θ_{opt} can be specified in terms of the ys

$$\theta_{\text{opt}} = 2\big(y_1^2 + y_2^2 + y_3^2 + y_4^2\big)^{\frac{1}{2}}\,. \tag{5.49}$$

Therefore, we have $y = 0$ for $\theta_{\text{opt}} = 0$. Then the constraints are quite transparent

$$w_{\text{cstr}}(x, 0) = 0 \;\text{ unless }\; \big(x_1^2 + x_2^2\big)^{\frac{1}{2}} + \big(x_3^2 + x_4^2\big)^{\frac{1}{2}} \leq 1 \tag{5.50}$$

The above special form reminds us the of following coordinate transformation:

$$\begin{aligned} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} \cos\varphi_1 & -\sin\varphi_1 \\ \sin\varphi_1 & \cos\varphi_1 \end{pmatrix} \begin{pmatrix} r_1 \\ 0 \end{pmatrix} = \begin{pmatrix} r_1\cos\varphi_1 \\ r_1\sin\varphi_1 \end{pmatrix}, \\ \begin{pmatrix} x_3 & x_4 \end{pmatrix} &= \begin{pmatrix} r_2 & 0 \end{pmatrix} \begin{pmatrix} \cos\varphi_2 & \sin\varphi_2 \\ -\sin\varphi_2 & \cos\varphi_2 \end{pmatrix} = \begin{pmatrix} r_2\cos\varphi_2 & r_2\sin\varphi_2 \end{pmatrix}, \\ (\mathrm{d}x) &= \mathrm{d}r_1\, r_1\, \mathrm{d}\varphi_1\, \mathrm{d}r_2\, r_2\, \mathrm{d}\varphi_2 \end{aligned} \tag{5.51}$$

[7] See [106] for a detailed account of the properties of two-qubit states and their classification.

with $0 \leq r_1 \leq 1 - r_2 \leq 1$ and $0 \leq \varphi_1, \varphi_2 \leq 2\pi$. These x values make up a four-dimensional volume

$$\int (\mathrm{d}x) = (2\pi)^2 \int_0^1 \mathrm{d}r_1\, r_1 \int_0^{1-r_1} \mathrm{d}r_2\, r_2 = \frac{\pi^2}{6} \tag{5.52}$$

but, since there is no volume in the four-dimensional y space, the set of probabilities with $\theta_{\mathrm{opt}} = 0$ has no eight-dimensional volume — it has no size.

The generic state in this set has $r_1 + r_2 < 1$ and full rank. A finite, if small, four-dimensional ball is then available for the y values. All y values on the three-dimensional surface of the ball have the same value of θ_{opt}, equal to the diameter of the ball. The volume of the ball is proportional to θ_{opt}^4 and, therefore, we have

$$P_0(\Theta_{\mathrm{opt}}) \propto \Theta_{\mathrm{opt}}^4 \quad \text{for} \quad 0 \lesssim \Theta_{\mathrm{opt}} \ll 1\,. \tag{5.53}$$

5.4.4.2 The vicinity of $\Theta_{\mathrm{opt}} = \sqrt{8}$

Different from the fact that we reach $\Theta = \sqrt{8}$ for only one state, we are able to obtain $\Theta_{\mathrm{opt}} = \sqrt{8}$ for all maximally entangled states with $\langle \sigma_y \otimes \sigma_y \rangle^2 = 1$. It means that $x = 0$, and the constraints are

$$w_{\mathrm{cstr}}(0, y) = 0 \text{ unless the two characteristic values of } \begin{pmatrix} y_1 & y_2 \\ y_3 & y_4 \end{pmatrix} \text{ are} \leq 1. \tag{5.54}$$

For convenience, we speicify the y matrix by the product of the rotation matrix and diagnal matrix

$$\begin{pmatrix} y_1 & y_2 \\ y_3 & y_4 \end{pmatrix} = \begin{pmatrix} \cos\phi_1 & -\sin\phi_1 \\ \sin\phi_1 & \cos\phi_1 \end{pmatrix} \begin{pmatrix} \vartheta_1 & 0 \\ 0 & \vartheta_2 \end{pmatrix} \begin{pmatrix} \cos\phi_2 & \sin\phi_2 \\ -\sin\phi_2 & \cos\phi_2 \end{pmatrix} \tag{5.55}$$

$$= \begin{pmatrix} \vartheta_1 \cos\phi_1 \cos\phi_2 + \vartheta_2 \sin\phi_1 \sin\phi_2 & \vartheta_1 \cos\phi_1 \sin\phi_2 - \vartheta_2 \sin\phi_1 \cos\phi_2 \\ \vartheta_1 \sin\phi_1 \cos\phi_2 - \vartheta_2 \cos\phi_1 \sin\phi_2 & \vartheta_1 \sin\phi_1 \sin\phi_2 + \vartheta_2 \cos\phi_1 \cos\phi_2 \end{pmatrix}$$

with $0 \le \vartheta_1 \le 1$, $-1 \le \vartheta_2 \le 1$, $0 \le \phi_1, \phi_2 \le 2\pi$, where ϑ_1 and $|\vartheta_2|$ are the characteristic values. The determinant $\vartheta_1\vartheta_2$ can be positive or negative; we avoid double coverage by restricting ϑ_1 to positive values while letting ϕ_1 and ϕ_2 range over a full 2π period.

After the switching from y-coordinate to ϕ-coordinate, the corresponding Jacobian factor in

$$(\mathrm{d}y) = \mathrm{d}\vartheta_1\,\mathrm{d}\vartheta_2\,\mathrm{d}\phi_1\,\mathrm{d}\phi_2\,|\vartheta_1^2 - \vartheta_2^2| \tag{5.56}$$

vanishes when $\vartheta_1 = |\vartheta_2| = 1$ and $\Theta_{\text{opt}} = 2(\vartheta_1^2 + \vartheta_2^2)^{\frac{1}{2}} = \sqrt{8}$. Therefore, there is no nonzero four-dimensional volume in the y space for $\Theta_{\text{opt}} = \sqrt{8}$. More specifically, the y-space volume for $\vartheta_1^2 + \vartheta_2^2 > \frac{1}{4}\Theta_{\text{opt}}^2$ is

$$\begin{aligned}
& (2\pi)^2 \int_0^1 \mathrm{d}\vartheta_1 \int_{-1}^1 \mathrm{d}\vartheta_2\,|\vartheta_1^2 - \vartheta_2^2|\,\eta\Big(4\big(\vartheta_1^2 + \vartheta_2^2\big) - \Theta_{\text{opt}}^2\Big) \\
= \;& (2\pi)^2 \left[\frac{2}{3} - \frac{1}{32}\Theta_{\text{opt}}^4 + \frac{1}{6}\big(\Theta_{\text{opt}}^2 - 4\big)^{\frac{3}{2}}\,\eta\big(\Theta_{\text{opt}}^2 - 4\big)\right] \\
= \;& \frac{\sqrt{8}\,\pi^3}{3}\Big(\sqrt{8} - \Theta_{\text{opt}}\Big)^3 + O\bigg(\Big(\sqrt{8} - \Theta_{\text{opt}}\Big)^4\bigg) \quad \text{for} \quad \Theta_{\text{opt}} \lesssim \sqrt{8}\,.
\end{aligned} \tag{5.57}$$

With respect to the corresponding x-space volume, we note that the maximally entangled states with

$$\begin{aligned}
\begin{pmatrix} y_1 & y_2 \\ y_3 & y_4 \end{pmatrix} &= \begin{pmatrix} \cos(\phi_1 - \phi_2) & \sin(\phi_1 - \phi_2) \\ -\sin(\phi_1 - \phi_2) & \cos(\phi_1 - \phi_2) \end{pmatrix} \\
\text{or} \qquad \begin{pmatrix} y_1 & y_2 \\ y_3 & y_4 \end{pmatrix} &= \begin{pmatrix} \cos(\phi_1 + \phi_2) & \sin(\phi_1 + \phi_2) \\ \sin(\phi_1 + \phi_2) & -\cos(\phi_1 + \phi_2) \end{pmatrix}
\end{aligned} \tag{5.58}$$

are equivalent because local unitary transformations turn them into each other. It is, therefore, sufficient to consider an ϵ-vicinity of one such state, for which we take that with $y_1 = y_4 = -1$ and $y_2 = y_3 = 0$. This is $|+\rangle\langle+|$ of Eq. (5.33), with ρ_ϵ in Eq. (5.37).

As a consequence of Eq. (5.34), we have

$$x_1 + x_3 \propto \epsilon^2\,,\;\; x_1 - x_3 \propto \epsilon \quad \text{and} \quad x_2 + x_4 \propto \epsilon^2\,,\;\; x_2 - x_4 \propto \epsilon\,, \tag{5.59}$$

so that the x-space volume is proportional to ϵ^6. Since we know from Eq. (5.43) that $\sqrt{8} - \Theta_{\text{opt}} \propto \epsilon^2$, it follows that the x-space volume is proportional to $\left(\sqrt{8} - \Theta_{\text{opt}}\right)^3$. Together with the y-space volume in Eq. (5.57), we so find that

$$1 - P_0(\Theta_{\text{opt}}) \propto \left(\sqrt{8} - \Theta_{\text{opt}}\right)^6 \quad \text{for} \quad 0 \lesssim \sqrt{8} - \Theta_{\text{opt}} \ll 1\,. \tag{5.60}$$

5.4.4.3 Analog of Eq. (5.31) and Eq. (5.32) for $P_0(\Theta_{\text{opt}})$

Since the absence of the symmetry for $P_0(\Theta_{\text{opt}})$, i.e., $W(\Theta_{\text{opt}}) \neq W(\sqrt{8} - \Theta_{\text{opt}})$, we use $B_{\alpha,\beta}(\Theta_{\text{opt}})$ instead

$$B_{\alpha,\beta}(\Theta_{\text{opt}}) = \left(\frac{1}{8}\right)^{\frac{1}{2}(\alpha+\beta+1)} \frac{(\alpha+\beta+1)!}{\alpha!\;\beta!} \int_0^{\Theta_{\text{opt}}} \mathrm{d}x\, x^\alpha (\sqrt{8} - x)^\beta\,. \tag{5.61}$$

as the ingredient of the approximation function

$$P_0(\Theta_{\text{opt}}) \simeq P_0^{(0)}(\Theta_{\text{opt}}) = \sum_l w_l B_{\alpha_l,\beta_l}(\Theta_{\text{opt}}) \tag{5.62}$$

with $\sum_l w_l = 1$.

For the corresponding approximation for $W_0(\Theta_{\text{opt}})$, which is the derivative of prior content $W_0(\Theta_{\text{opt}}) = \dfrac{\mathrm{d}}{\mathrm{d}\Theta_{\text{opt}}} P_0(\Theta_{\text{opt}})$, the fitting parameters are: One of the powers α_l is equal to 3 (see Eq. (5.53)) and one of the β_ls is equal to 5 (see Eq. (5.53)), and the other ones are larger. For the sample of 500 000 sets of probabilities that generated the red Θ_{opt} histograms in Fig. 5.5(left), a fit with a mean squared error of 4.2×10^{-4} is achieved by a five-term approximation with

these parameter values:

l	w_l	α_l	β_l
1	0.2187	3	5.2467
2	0.2469	5.2238	5
3	0.3153	14.1703	11.7922
4	0.2478	7.9878	11.8061
5	−0.0287	37.5270	15.7518

(5.63)

There are 12 fitting parameters here. The black curve to that histogram shows the corresponding approximation for $W_0(\Theta_{\text{opt}}) = \dfrac{\mathrm{d}}{\mathrm{d}\Theta_{\text{opt}}} P_0(\Theta_{\text{opt}})$. See Figs. 5.5 and 5.8 for illustrations.

5.4.5 Iterated MC integrations for $P_0(\Theta_{\text{opt}})$

Analogous to Sec. 5.4.3, we show the subsequent approximations $P_0^{(1)}(\Theta_{\text{opt}})$, $P_0^{(2)}(\Theta_{\text{opt}})$, and $P_0^{(3)}(\Theta_{\text{opt}})$ in Fig. 5.8(left) indicated by the blue, cyan and red dots, respectively, so does the points after subtracting $\Theta_{\text{opt}}/\sqrt{8}$ in Fig. 5.8(right). Similarly, the low-pass filter is used to removes the noise, such that the resulting truncated Fourier series gives a more smoothing curve.

5.4.6 Likelihood and optimal error regions

In each iteration round, we obtain an updated approximation $\widetilde{W}_{\text{r},0}^{(n)}(F)$. We sample the probability space in accordance with $w_0^{(n)}(p)L(D|p)$ for a MC integration of the posterior density. Here, the updated prior density is

$$w_0^{(n)}(p) = \frac{w_{\text{r}}(p)}{\widetilde{W}_{\text{r},0}^{(n)}(f(p))}. \tag{5.64}$$

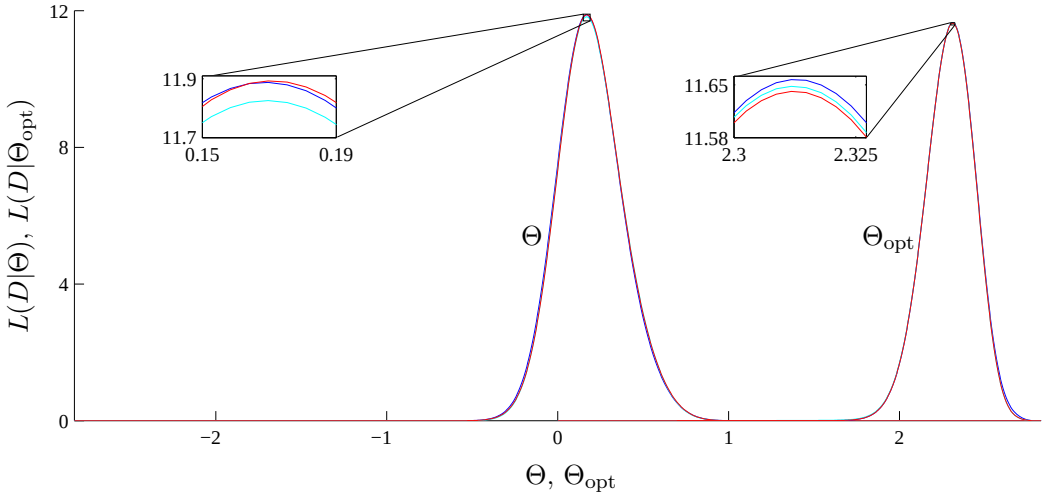

Figure 5.9: Likelihood function for Θ and Θ_{opt}. The plot of $L(D|\Theta)$ shows the Θ-likelihood obtained for the three subsequent iterations in Fig. 5.6(b), with a blow-up of the region near the maximum. The colors blue, cyan, and red correspond to those in Fig. 5.6. — The plot of $L(D|\Theta_{\mathrm{opt}})$ is analogous; it shows the likelihoods for iterations $n = 0$, 1, and 2 in green, blue, and cyan, respectively.

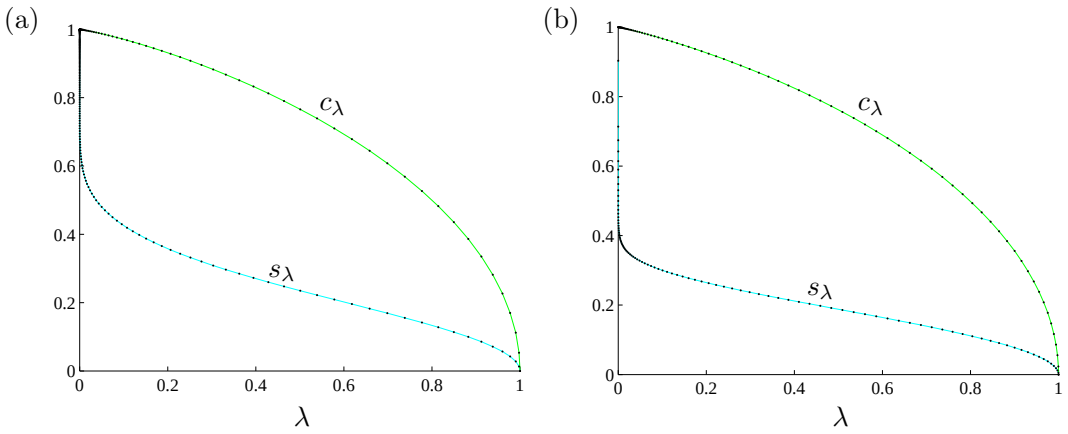

Figure 5.10: Size and credibility of bounded-likelihood intervals for the CHSH quantities, computed from the likelihood functions in Fig. 5.9. (a) Fixed measurement of Eq. (5.24) with the primitive prior of Eq. (3.11); (b) optimized measurement of Eq. (5.26) with the Jeffreys prior of Eq. (3.12).

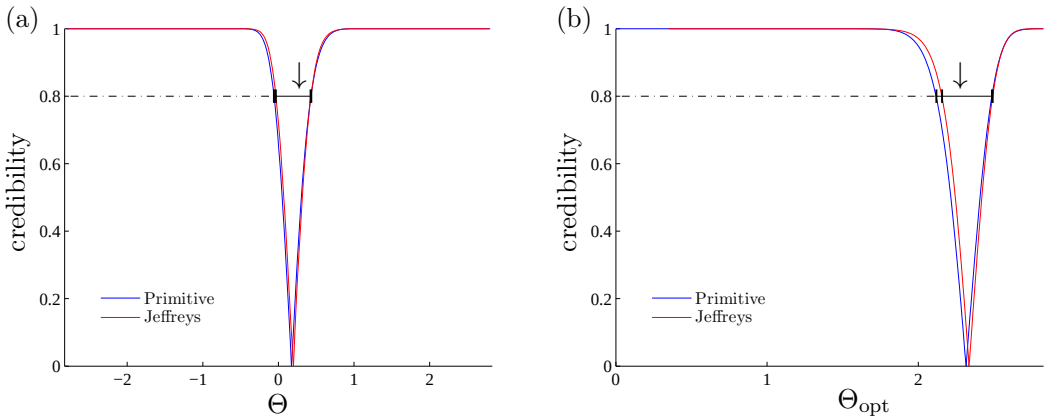

Figure 5.11: Optimal error intervals for (a) Θ and (b) Θ_{opt}. The blue and red curves delineate the boundaries of the SCIs in the same manner as in Figs. 5.3 and 5.4. Both true values are inside the indicated SCRs with credibility 0.8.

Then the corresponding $P_D^{(n)}(\Theta)$ and $W_D^{(n)}(\Theta)$ are obtained. With $W_0^{(n)}(\Theta)$ and $W_D^{(n)}(\Theta)$ at hand, we can determine the Θ-likelihood $L(D|\Theta)$ immediately; and likewise for $L(D|\Theta_{\text{opt}})$. Figure 5.9 shows the sequence of approximations.

Note that the approximations for the Θ-likelihood hardly change from one iteration to the next, so that just a few rounds are enough and we quit the iterations. With the $L(D|\Theta)$ and $L(D|\Theta_{\text{opt}})$ at hand, we proceed to the calculation of the size s_λ and the credibility c_λ of the BLIs, which are shown in Fig. 5.10 for Θ and the primitive prior as well as Θ_{opt} and the Jeffreys prior. The plots for the respective other prior are very similar.

Analogous to the situations in Sec. 5.3, there is not much difference in the SCIs obtained for the two priors, although the number of measured copies ($N = 180$) is not large; see Fig. 5.11. Another observation, as we mentioned in the preceding section, is that the advantage of Θ_{opt} over Θ is obvious: Whereas virtually all Θ-SCIs with non-unit credibility are inside the range $-2 < \Theta < 2$, the indicated Θ_{opt}-SCI with credibility 0.8 is entirely in the range $\Theta_{\text{opt}} > 2$.

5.5 Sampling error analysis

To evaluate the high-dimensional integrals (see Eqs. (5.1) and (5.2)), Monte Carlo methods are adapted. Since the size of sample is finite, the statistics, say $P_{\mathrm{r},0}(F)$, on the sample is generally different from the true value. These discrepancies between the sample and the true value are considered as the sampling error. More broadly, sampling error refers to the phenomenon of random sampling variation. Sampling error is a measure for judging the quality of the random sampling. In this section, we take the error analysis of $P_{\mathrm{r},0}(F)$ as example (see Eq. (5.1)). The primitive prior is employed for the integral of $P_{\mathrm{r},0}(F)$. We explore how the sampling error of $P_{\mathrm{r},0}(F)$ changes as the sample size N varies.

We consider the relative error

$$\delta x = \frac{\Delta x}{x} = \frac{x_0 - x}{x}, \tag{5.65}$$

as a measure of sampling error, where x is the true value and x_0 is the estimated value. In practice, however, the true value of parameter x is usually unknown. Therefore, we substitute the standard deviation σ for absolute error Δx, and the mean $\bar{x}$ for x. After the substitution, the relative error is

$$\delta x = \frac{\sigma}{\bar{x}}. \tag{5.66}$$

Roughly speaking, as the number of sampling points N increases, the sampling error will decrease. For a certain value of CHSH Θ, it is easy to prove that the relationship between relative error $\delta P_{\mathrm{r},0}(\Theta)$ and the sample size N is

$$\log \delta P_{\mathrm{r},0}(\Theta) = \log \sqrt{\frac{1 - \overline{P_{\mathrm{r},0}(\Theta)}}{\overline{P_{\mathrm{r},0}(\Theta)}}} - \frac{1}{2} \log N. \tag{5.67}$$

Therefore, the logarithm of relative error $\log \delta P_{\mathrm{r},0}(\Theta)$ is a linear function of $\log N$

with the slope being $-\frac{1}{2}$, and the intercept is a function of $\overline{P_{\mathrm{r},0}(\Theta)}$.

We choose three CHSH values $\Theta = \{-2, 0, 2\}$ as examples. For the MC integration, the sample size is varied from $N = 400$ to $N = 400\,000$. For each N, we sampling the probability space 100 times with the unweighted sample (see footnote 6). For each time, certain value of $P_{\mathrm{r},0}(\Theta)$ is evaluated by a numerical integral. With 100 values of $P_{\mathrm{r},0}(\Theta)$, the standard error σ and mean $\overline{P_{\mathrm{r},0}}(\Theta)$ are obtained.

In Fig. 5.12, we show that the logarithm of the relative error $\log \delta P_{\mathrm{r},0}(\Theta)$ is a linear function of $\log N$ for three values of CHSH. The solid lines specify the theoretical predictions in Eq. (5.67). As a particular case, $P_{\mathrm{r},0}(\Theta = 0) = \frac{1}{2}$ is known because of the symmetry $P_{\mathrm{r},0}(\Theta) + P_{\mathrm{r},0}(-\Theta) = 1$. For the other two values $\Theta = \{2, -2\}$, we use the MC integration values $\widetilde{P_{\mathrm{r},0}}(\Theta)$ instead. The points are obtained using the unweighted samples obtained by bootstrapping the weighted sample. Statistical noise aside, the MC integration values are consistent with the theoretical prediction. Therefore, the sample used for the MC integration is reliable.

5.6 Summary

Analogous to the optimal error regions for quantum state estimation, we first introduce the notions of the size and credibility of a range of the property of interest in the context of state-property estimation. Then we propose maximum-likelihood intervals and smallest credible intervals as the optimal error intervals for state-property estimation. We prove that both of them are the bounded-likelihood intervals, where the likelihood here is the F-likelihood $L(D|F)$— the likelihood conditional on the property F.

The F-likelihood $L(D|F)$ is central for the construction of optimal error intervals for state-property estimation. However, it is usually difficult to compute

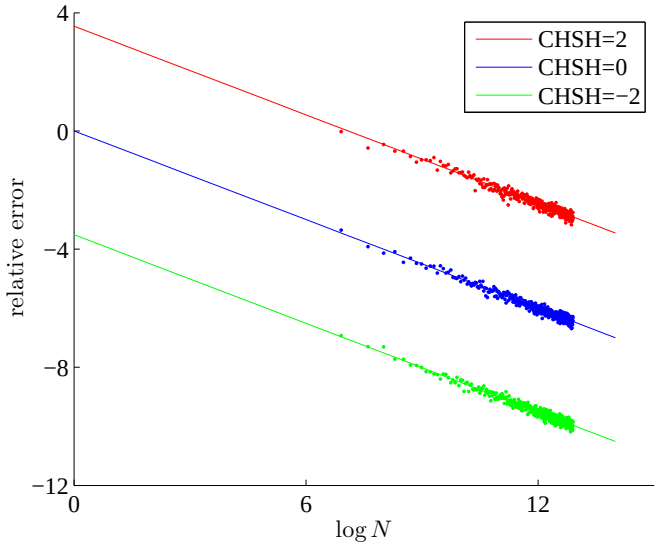

Figure 5.12: The logarithm of the relative error $\log \delta P_{\mathrm{r},0}(F)$ as a linear function of $\log N$ for three values of CHSH. Solid lines represent the linear function Eq. (5.67), while the points are obtained using the MC integration. Statistical noise aside, the MC integration values are consistent with the theoretical prediction.

$L(D|F)$ analytically, except for the case of single-qubit state property. Therefore, we integrate the high-dimensional integrals with MC techniques. Since the size of sample is finite, the MC integration is not precise enough to distinguish the values in finite intervals near the boundaries. This problem can be solved by the iteration algorithm we provide and the known power law near the boundary. Then, the corresponding F-likelihood is reliable for all values of F.

To illustrate of the algorithm, we estimate the properties of single-qubit state, e.g., the fidelity with respect to target state and purity as examples, as well as the CHSH quantity for two-qubit states.

Chapter 6

Conclusion and Outlook

Quantum state estimation is central to many tasks of quantum information processing, such as quantum communication, quantum computation and quantum cryptography. Two branches of quantum state estimation get much attention: the schemes of designing measurements that can increase the efficiency of the estimation, and the data processing methods which provide a reliable and optimal estimator for the quantum state. As for the data processing methods, point estimators and regions estimators are complement to each other. For point estimators, many methods have been proposed, such as linear inversion, maximum likelihood estimation, maximum entropy estimation and so on. Each method has its pros and cons. While for region estimators, several proposals are devised based on frequentist methodology and Bayesian strategy.

Our method for constructing the optimal error regions for quantum state estimation is described in Chapter 3. We measure the size of a region by its prior content, and propose maximum-likelihood regions and smallest credible regions as the optimal error regions. Then, we prove that these optimal error regions are bounded-likelihood regions, and illustrate the method by simulated single-qubit and two-qubit experiments.

In this thesis, we mainly discuss the direct estimation for properties of the quantum state in Chapters 4 and 5. Usually, we are more interested a few functions of the quantum state than the full details of the quantum system. Therefore, a direct estimate of the properties of the state is more efficient and

more meaningful than the indirect estimation method in which one first estimates the quantum state then take the value of function of the corresponding state to be the guess for that property. We study a single function of the state and construct the optimal error intervals as the optimal error intervals for state-property estimation. Analogous to optimal error regions for quantum state estimation, we propose maximum-likelihood intervals and smallest credible intervals for the state-property estimation. For illustration, we study the fidelity (with respect to target state) and purity of single-qubit states and the CHSH quantity for two-qubit states. The extension of this method for estimating the new property and to estimate multiple properties simultaneously requires further study. In addition, this method might be used to estimate the parameters which parameterize the quantum process and quantum measurements. It should be noticed that, to construct error regions for high-dimensional quantum system, a smarter numerical integration algorithm is required for the multiple dimension integral.

A major problem left open is how to infer the quantum state and properties of the state as efficiently as possible. The dimension of the Hilbert space increases exponentially with the size of the n-qubit system. To perform a full quantum state estimation for such a system, the number of measurements required are extremely large. In addition, a direct estimate of the degree of entanglement of high-dimensional quantum system is usually difficult. Therefore, efficient schemes of measurement designing and smarter data processing protocols require further study for solving these problems .

Appendix A

Derivation of Eqs. (3.28) and (3.29)

From Eqs. (3.26) and (3.27), the derivative of c_λ with respect to λ is

$$\begin{aligned}
\frac{\partial c_\lambda}{\partial \lambda} &= \frac{1}{L(D)S_0}\int_{\mathcal{R}_0}(\mathrm{d}\rho)L(D|p)\,\frac{\partial \chi_\lambda(p)}{\partial \lambda} \\
&= \frac{1}{L(D)S_0}\int_{\mathcal{R}_0}(\mathrm{d}\rho)L(D|p)\frac{\partial}{\partial \lambda}\eta\Big(L(D|p)-\lambda L(D|\widehat{p}_{\mathrm{ML}})\Big) \\
&= \frac{1}{L(D)S_0}\int_{\mathcal{R}_0}(\mathrm{d}\rho)L(D|p)\delta\Big(L(D|p)-\lambda L(D|\widehat{p}_{\mathrm{ML}})\Big) \\
&= \frac{\lambda L(D|\widehat{p}_{\mathrm{ML}})}{L(D)S_0}\int_{\mathcal{R}_0}(\mathrm{d}\rho)\frac{\partial \chi_\lambda(p)}{\partial \lambda} \\
&= \frac{\lambda L(D|\widehat{p}_{\mathrm{ML}})}{L(D)}\frac{\partial s_\lambda}{\partial \lambda}
\end{aligned} \tag{A.1}$$

Therefore, we have the Eq. (3.28):

$$L(D)\frac{\partial}{\partial \lambda}c_\lambda = L(D|\widehat{p}_{\mathrm{ML}})\,\lambda\frac{\partial}{\partial \lambda}s_\lambda\,. \tag{A.2}$$

From Eq. (A.1), we integrate both side of the equation from 1 to λ' and obtain

$$\int_1^{\lambda'}\mathrm{d}\lambda\frac{\partial}{\partial \lambda}c_\lambda = \frac{L(D|\widehat{p}_{\mathrm{ML}})}{L(D)}\int_1^{\lambda'}\mathrm{d}\lambda\left(\lambda\frac{\partial}{\partial \lambda}s_\lambda\right)\,. \tag{A.3}$$

The left hand side is $c_{\lambda'}-c_1=c_{\lambda'}$, because of the fact that $c_1=0$. Using the

integration by parts, the right hand side is

$$\begin{aligned} & \frac{L(D|\widehat{p}_{\mathrm{ML}})}{L(D)} \int_1^{\lambda'} \mathrm{d}\lambda \left[\frac{\partial}{\partial \lambda} (\lambda s_\lambda) - s_\lambda \right] \\ = \; & \frac{L(D|\widehat{p}_{\mathrm{ML}})}{L(D)} \left(\lambda' s'_\lambda + \int_{\lambda'}^1 \mathrm{d}\lambda \, s_\lambda \right) \end{aligned} \tag{A.4}$$

Compare Eqs. (A.4) and (3.29), the only work left is to prove that $L(D)/L(D|\widehat{p}_{\mathrm{ML}}) = \int_0^1 \mathrm{d}\lambda \; s_\lambda$. The proof is in the following:

$$L(D) = \frac{1}{S_0} \int_{\mathcal{R}_0} (\mathrm{d}\rho) \, L(D|p) = L(D|\widehat{p}_{\mathrm{ML}}) \int_{\mathcal{R}_0} (\mathrm{d}\rho) \frac{L(D|p)}{L(D|\widehat{p}_{\mathrm{ML}})} \, . \tag{A.5}$$

By using the property of the step function, we translate $L(D|p)/L(D|\widehat{p}_{\mathrm{ML}})$ into the integral of the step function (note that $0 \leq L(D|p)/L(D|\widehat{p}_{\mathrm{ML}}) \leq 1$):

$$\frac{L(D|p)}{L(D|\widehat{p}_{\mathrm{ML}})} = \int_0^1 \mathrm{d}\lambda \; \eta \left(\frac{L(D|p)}{L(D|\widehat{p}_{\mathrm{ML}})} - \lambda \right) \tag{A.6}$$

Substituting Eq. (A.6) into Eq. (A.5), we obtain

$$\begin{aligned} L(D) &= L(D|\widehat{p}_{\mathrm{ML}}) \int_{\mathcal{R}_0} (\mathrm{d}\rho) \int_0^1 \mathrm{d}\lambda \; \eta \left(\frac{L(D|p)}{L(D|\widehat{p}_{\mathrm{ML}})} - \lambda \right) \\ &= L(D|\widehat{p}_{\mathrm{ML}}) \int_0^1 \mathrm{d}\lambda \int_{\mathcal{R}_0} (\mathrm{d}\rho) \eta \left(L(D|p) - \lambda L(D|\widehat{p}_{\mathrm{ML}}) \right) \\ &= L(D|\widehat{p}_{\mathrm{ML}}) \int_0^1 \mathrm{d}\lambda s_\lambda \end{aligned} \tag{A.7}$$

Accordingly, combining Eqs. (A.3), (A.4) and (A.7), and interchanging λ and λ', we obtain Eq. (3.29).

Bibliography

[1] J. Shang, H. K. Ng, A. Sehrawat, X. Li, and B.-G. Englert. Optimal error regions for quantum state estimation. *New. J. Phys.*, 15:123026, 2013.

[2] C. W. Helstrom. *Quantum Detection and Estimation Theory.* Academic Press, New York, 1976.

[3] M. G. A. Paris and J. Řeháček (Eds.). *Quantum State Estimation*, volume 649 of *Lecture Notes in Physics.* Springer-Verlag, Berlin Heidelberg, 2004.

[4] K. Vogel and H.Risken. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Phys. Rev. A*, 40:2847, 1989.

[5] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani. Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacumm. *Phys. Rev. Lett.*, 70:1244, 1993.

[6] K. Banaszek, C. Radzewic, K. Wodkiewicz, and J. S. Krasinski. Direct measurement of the Wigner function by photon counting. *Phys. Rev. A*, 60:674, 1999.

[7] K. Banaszek, G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi. Maximum-likelihood estimation of the density matrix. *Phys. Rev. A*, 61:010304(R), 2000.

[8] A. G. White, D. F. V. James, W. J. Munro, and P. G. Kwiat. Exploring Hillbert space: Accurate characterization of quantum information. *Phys. Rev. A*, 65:012301, 2002.

[9] A. Ourjoumtsev, H. Jeong, R. Tualle-Brouri, and P. Grangier. Generation of optical 'Schrödinger cats' from photon number states. *Nature*, 448:784, 2007.

[10] J. S. Neergaard-Nielsen, B. Melholt Nielson, C. Hettich, K. Mølmer, and E. S. Polzik. Generation of a superposition of odd photon number states for quantum information networks. *Phys. Rev. Lett.*, 97:083604, 2006.

[11] J. Řeháček, B.-G. Englert, and D. Kaszlikowski. Minimal qubit tomography. *Phys. Rev. A*, 70:052321, 2004.

[12] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2010.

[13] I. L. Chuang and M. A. Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *J. Mod. Opt*, 44:2455, 1997.

[14] J. F. Poyatos, J. I. Cirac, and P. Zoller. Complete characterization of a quantum process: The two-bit quantum gate. *Phys. Rev. Lett.*, 78:390, 2010.

[15] J. B. Altepeter, D. Branning, E. Jeffrey, T. C. Wei, P. G. Kwiat, R. T. Thew, J. L. O'Brien, M. A. Nielsen, and A. G. White. Ancilla-assisted quantum process tomography. *Phys. Rev. Lett.*, 90:193601, 2003.

[16] G. M. D'Ariano and L. Maccone. Measuring quantum optical Hamiltonians. *Phys. Rev. Lett.*, 80:5465, 1998.

[17] M. A. Nielsen, E. Knill, and R. Laflamme. Complete quantum teleportation using nuclear magnetic resonance. *Nature*, 396:52, 1998.

[18] M. W. Mitchell, C. W. Ellenor, S. Schneider, and A. M. Steinberg. Diagnosis, prescription and prognosis of a Bell-state filter by quantum process tomography. *Phys. Rev. Lett.*, 91:120402, 2003.

[19] A. Luis and L. L. Sanchez-Soto. Complete characterization of arbitrary quantum measurement processes. *Phys. Rev. Lett.*, 83:3573, 1999.

[20] J. Fiurasek. Maximum-likelihood estimation of quantum measurement. *Phys. Rev. A*, 64:024102, 2001.

[21] G. M. D'Ariano, L. Maccone, and P. Lo Presti. Quantum calibration of measurement instrumentation. *Phys. Rev. Lett.*, 93:250407, 2004.

[22] J. S. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. L. Pregnell, Ch. Silberhorn, T. C. Ralph, J. Eisert, M. B. Plenio, and I. A.Walmsley. Tomography of quantum detectors. *Nature Physics*, 5:27, 2008.

[23] R. L. Plackett. The principle of the arithmetic mean. *Biometrika*, 45:130, 1958.

[24] G. A. Barnard. Thomas Bayes essay towards solving a problem in the doctrine of chances. *Biometrika*, 45:293, 1958.

[25] R. L. Plackett. A historical note on the method of least-squares. *Biometrika*, 36:458, 1949.

[26] I. Todhunter. *A history of the mathematical theory of probability*. Macmillan, New York, 1865.

[27] R. A. Fisher. On an absolute criterion for fitting frequency curves. *Mess. Math*, 41:155, 1912.

[28] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc. Lond. A*, 222:309, 1922.

[29] R. A. Fisher. Theory of statistical estimation. *Math. Proc. Cambr. Phil. Soc.*, 22:700, 1925.

[30] W. Pauli. *General Principles of Quantum Mechanics*. Springer-Verlag, Berlin, 1990.

[31] U. Fano. Description of states in quantum mechanics by density matrix and operator techniques. *Rev. Mod. Phys.*, 29:74–93, 1957.

[32] W. Gale, E. Guth, and G. T. Trammell. Determination of the quantum state by measurements. *Phys. Rev.*, 165:1434–1436, 1968.

[33] R. G. Newton and B.-L. Young. Measurability of the spin density matrix. *Ann. Phys.*, 49:393, 1968.

[34] W. Band and J. Park. The empirical determination of quantum states. *Found. Phys.*, 1:133–144, 1970.

[35] J. Park and W. Band. A general method of empirical state determination in quantum physics: Part I. *Found. Phys.*, 1:211–216, 1971.

[36] W. Band and J. Park. A general method of empirical state determination in quantum physics: Part II. *Phys. Rev. A*, 1:339–357, 1971.

[37] A. I. Lvovsky and M. G. Raymer. Continuous-variable optical quantum-state tomography. *Rev. Mod. Phys.*, 81:299, 2009.

[38] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, 1957.

[39] E. T. Jaynes. Information theory and statistical mechanics. II. *Phys. Rev.*, 108:171–190, 1957.

[40] V. Bužek, G. Drobný, G. Adam, R. Derka, and P. L. Knight. Reconstruction of quantum states of spin systems via the Jaynes principle of maximum entropy. *J. Mod. Opt.*, 44(11-12):2607–2627, 1997.

[41] V. Bužek, G. Drobný, R. Derka, G. Adam, and H. Wiedemann. Quantum state reconstruction from incomplete data. *Chaos, Solitons Fractals*, 10(6):981–1074, 1999.

[42] R. Blume-Kohout. Hedged maximum likelihood quantum state estimation. *Phys. Rev. Lett.*, 105:200504, 2010.

[43] B. P. Carlin and T. A. Louis. *Bayesian Methods for Data Analysis (Texts in Statistical Science)*. 3rd ed., Chapman & Hall/CRC, New York, 2008.

[44] K. R. W. Jones. Principles of quantum inference. *Ann. Phys.*, 207:140–170, 1991.

[45] R. Schack, T. A. Brun, and C. M. Caves. Quantum Bayes rule. *Phys. Rev. A*, 64:014305, 2001.

[46] F. Neri. Quantum Bayesian methods and subsequent measurements. *Phys. Rev. A*, 72:062306, 2005.

[47] F. Tanaka and F. Komaki. Bayesian predictive density operators for exchangeable quantum-statistical models. *Phys. Rev. A*, 71:052323, 2005.

[48] R. Blume-Kohout. Optimal, reliable estimation of quantum states. *New J. Phys.*, 12:043034, 2010.

[49] G. M. D'Ariano, M. F. Sacchi, and J. Kahn. Minimax quantum-state discrimination. *Phys. Rev. A*, 72:032310, 2005.

[50] G. M. D'Ariano, M. F. Sacchi, and J. Kahn. Minimax discrimination of two Pauli channels. *Phys. Rev. A*, 72:052302, 2005.

[51] M. Guţă and L. Artiles. Minimax estimation of the Wigner function in quantum homodyne tomography with ideal detectors. *Math. Meth. Stat.*, 16:1–15, 2007.

[52] H. K. Ng and B.-G. Englert. A simple minimax estimator for quantum states. *Int. J. Quantum Inf.*, 10:1250038, 2012.

[53] H. K. Ng, K. T. B. Phuah, , and B.-G. Englert. Minimax mean estimator for the trine. *New J. Phys.*, 14:085007, 2012.

[54] D. W. Leung. *Towards Robust Quantum Computation.* PhD thesis, Stanford University, 2000.

[55] D. W. Leung. Choi's proof as a recipe for quantum process tomography. *J. Math. Phys.*, 44:528–533, 2003.

[56] G. M. D'Ariano, P. Lo Presti, and M. F. Sacchi. Bell measurements and observables. *Phys. Lett. A*, 272:32, 2000.

[57] J. B. Altepeter, D. Branning, E. Jeffrey, T. C. Wei, P. G. Kwiat, R. T. Thew, J. L. DBrien, M. A. Nielsen, and A. G. White. Ancilla-assisted quantum process tomography. *Phys. Rev. Lett.*, 90:193601, 2003.

[58] F. De Martini, A. Mazzei, M. Ricci, and G. M. D'Ariano. Exploiting quantum parallelism of entanglement for a complete experimental quantum characterization of a single-qubit device. *Phys. Rev. A*, 67:062307, 2003.

[59] M. Mohseni and D. A. Lidar. Direct characterization of quantum dynamics. *Phys. Rev. Lett.*, 97:170501, 2006.

[60] M. Mohseni and D. A. Lidar. Direct characterization of quantum dynamics: General theory. *Phys. Rev. A*, 75:062331, 2007.

[61] Z.-W. Wang, Y.-S. Zhang, Y.-F. Huang, X.-F. Ren, and G.-C. Guo. Experimental realization of direct characterization of quantum dynamics. *Phys. Rev. A*, 75:044304, 2007.

[62] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laflamme. Symmetrized characterization of noisy quantum processes. *Science*, 317:1893–1896, 2007.

[63] M. Mohseni, A. T. Rezakhani, and D. A. Lidar. Quantum-process tomography: Resource analysis of different strategies. *Phys. Rev. A*, 77:032322, 2008.

[64] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, 2009.

[65] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, 1935.

[66] E. Schrödinger. Can quantum-mechanical description of physical reality be considered complete? *Die Naturwissenschaften*, 23:807, 1935.

[67] R. Raussendorf and H. J. Briegel. A one-way quantum computer. *Phys. Rev. Lett.*, 86:5188–5191, 2001.

[68] A. Ekert. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.*, 67:661–663, 1991.

[69] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, 1993.

[70] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Phys. Rev. Lett.*, 69:2881–2884, 1992.

[71] A. Peres. Neumark's theorem and quantum inseparability. *Found. Phys.*, 20:1441–1453, 1990.

[72] P. Busch. Informationally complete-sets of physical quantities. *Int. J. Theor. Phys.*, 30:1217, 1991.

[73] C. Schwemmer, L. Knips, D. Richart, H. Weinfurter, T. Moroder, M. Kleinmann, and O. Gühne. Systematic errors in current quantum state tomography tools. *Phys. Rev. Lett.*, 114:080403, 2015.

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

[75] L. Chen, H. Zhu, and T.-C. Wei. Connections of geometric measure of entanglement of pure symmetric states to quantum state estimation. *Phys. Rev. A*, 83:012305, 2011.

[76] Y. S. Teo, H. Zhu, B.-G. Englert, J. Řeháček, and Z. Hradil. Quantum-state reconstruction by maximizing likelihood and entropy. *Phys. Rev. Lett.*, 107:020404, 2011.

[77] G. J. Lidstone. Note on the general case of the Bayes–Laplace formula for inductive or a posteriori probabilities. *Trans. Fac. Actuaries*, 8:182–192, 1920.

[78] Y. S. Teo. *Numerical Estimation Schemes for Quantum Tomography*. Ph.D. thesis, National University of Singapore, arXiv:1302:3399 [quant-ph], 2012.

[79] J. Řeháček J, D. Mogilevtsev, and Z. Hradil. Tomography for quantum diagnostics. *New J. Phys.*, 10:043022, 2008.

[80] K. M. R. Audenaert and S. Scheel. Quantum tomographic reconstruction with error bars: a Kalman filter approach. *New J. Phys.*, 11:023028, 2009.

[81] B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. New York: Chapman & Hall/CRC, 1993.

[82] M. Christandl and R. Renner. Reliable quantum state tomography. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 109:120403, 2012.

[83] R. Blume-Kohout. Robust error bars for quantum tomography. *arXiv:1202.5270 [quant-ph]*, 109:120403, 2012.

[84] S. Jiang. *Symmetric Minimal Quantum Tomography and Optimal Error Regions*. PhD thesis, National University of Singaore, 2014.

[85] M. J. Evans, I. Guttman, and T. Swartz. Optimality and computations for relative surprise inferences. *Can. J. Stat.*, 34:113, 2006.

[86] L. A. Wasserman. A robust Bayesian interpretation of likelihood regions. *Ann. Stat.*, 17:1387, 1989.

[87] Dai J, Teo Y S, Tan W H, Len Y L, Ng H K, and Englert B-G. *in preparation*.

[88] J. Y. Sim. Self-calibrating quantum state estimation. *BSc thesis, National Univesity of Singapore*, 2015.

[89] V. Giovannetti, S. Lloyd, and L. Maccone. Quantum-enhanced measurements: Beating the standard quantum limit. *Science*, 306:1330, 2004.

[90] V. Giovannetti, S. Lloyd, and L. Maccone. Advances in quantum metrology. *Nature Photonics*, 5:222, 2011.

[91] A. K. Ekert, C. M. Alves, D. K. L. Oi, M. Horodecki, P. Horodecki, and L. C. Kwek. Direct estimations of linear and nonlinear funcionals of a quantum state. *Phys. Rev. Lett*, 88:217901, 2002.

[92] P. Horodecki and A. K. Ekert. Method for direct detection of quantum entanglement. *Phys. Rev. Lett*, 89:127902, 2002.

[93] F. A. Bovino, G. Castagnoli G, A. K. Ekert, P. Horodecki, C. M. Alves, and A. V. Sergienko. Direct measurement of nonlinesar properties of bipartite quantum states. *Phys. Rev. Lett*, 95:240407, 2005.

[94] R. D. Somma, J. Chiaverini, and D. J. Berkeland. Lower bounds for the fidelity of entangled-state preparation. *Phys. Rev. A*, 74:052302, 2006.

[95] O. Gühne, C.-Y. Lu, W.-B. Gao, and J.-W. Pan. Toolbox for entanglement detection and fidelity estimation. *Phys. Rev. A*, 76:030305(R), 2007.

[96] R. Blume-Kohout, J. O. S. Yin, and S. J. van Enk. Entanglement verification with finite data. *Phys. Rev. Lett.*, 105:170501, 2010.

[97] S. T. Flammia and Y.-K. Liu. Direct fidelity estimation from few Pauli measurements. *Phys. Rev. Lett*, 106:230501, 2011.

[98] P. Faist P and R. Renner. Practical, reliable error bars in quantum tomography. *e-print arXiv 1509.06763[quant-ph]*, 2015.

[99] X. Li, J. Shang, H. K. Ng, and B.-G. Englert. Optimal error intervals for properties of the quantum state. *e-print arXiv 1602.05780[quant-ph]*, 2016.

[100] G. N. M. Tabia and B.-G. Englert. Efficient quantum key distribution with trines of reference-frame-free qubits. *Phys. Lett. A*, 375:817, 2011.

[101] J. Shang, Y.-L. Seah, H. K. Ng, D. J. Nott, and B.-G. Englert. Monte Carlo sampling from the quantum state space. I. *New. J. Phys.*, 17:043017, 2015.

[102] Y.-L. Seah, J. Shang, H. K. Ng, D. J. Nott, and B.-G. Englert. Monte Carlo sampling from the quantum state space. II. *New. J. Phys.*, 17:043018, 2015.

[103] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880, 1969.

[104] J. F. Clauser and M. A. Horne. Experimental consequences of objective local theories. *Phys. Rev. D*, 10:526, 1974.

[105] J. Shang, H. K. Ng, and B.-G. Englert. Quantum state tomography: Mean squared error matters, bias does not. *e-print arXiv:1405.5350[quant-ph]*, 2014.

[106] B.-G. Englert and N. Metwally. Kinematics of qubit pairs. chapter 2 in [107], 2002.

[107] G. Chen and R. K. Brylinski (eds). *Mathematics of Quantum Computation*. Boca Raton: Chapman and Hall, 2002.

Index