Optimal error regions for quantum state estimation

arXiv:1302.4081[quant-ph]; New J. Phys. **15** (2013) 123026

Jiangwei Shang[1], Hui Khoon Ng[1,2], Arun Sehrawat[1], Xikun Li[1], Berge Englert[1,3]

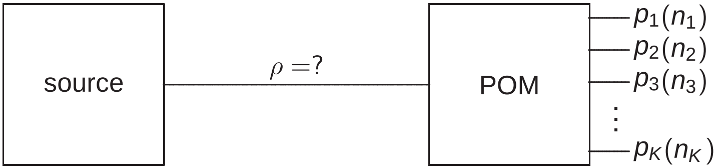
[1]Centre for Quantum Technologies (CQT), Singapore
[2]DSO National Laboratories, Singapore
[3]Department of Physics, National University of Singapore (NUS)

Centre For Quantum Technologies

DSO

NUS National University of Singapore

Scenario of quantum state estimation



The **source** emits independently and identically prepared quantum-information carriers whose relevant degrees of freedom are described by the "true" statistical operator ρ, which is unknown.

The **probability-operator measurement** (POM) has K outcomes Π_k that give rise to the "true" detection probabilities p_k in accordance with the Born rule, $p_k = \operatorname{tr}\{\rho\Pi_k\}$.

The **actual data** D consist of $n_1, n_2, \ldots, n_K$ detector clicks in *one particular sequence* upon measuring a total of $N = n_1 + n_2 + \cdots + n_K$ copies. [You may want to verify that the sequence is not untypical.]

State estimation: Exploit the data for an educated guess about $p = (p_1, p_2, \ldots, p_K)$; convert $p \to \rho$ if you can.

Principles of quantum state estimation

1 Be guided by common sense and the methods of classical statistical inference.*

2a Estimate event probabilities from the data, after measuring N copies.

2b Determine the estimator $\widehat{\rho}$ of the state from the estimated probabilities $\widehat{p}_1, \widehat{p}_2, \widehat{p}_3, \ldots$ and, if necessary, invoke additional criteria (such as Jaynes's maximum-entropy criterion).

Note 1: $n = (n_1, n_2, \ldots, n_K) \to \widehat{p} = (\widehat{p}_1, \ldots, \widehat{p}_K)$ is what the data tell us; $\widehat{p} \to \widehat{\rho}$ is often not unique, and then the data do *not* tell us $\widehat{\rho}$ and one needs those "additional criteria".

Note 2: $\widehat{p}_k \to p_k^{(\text{true})}$ for $N \to \infty$ ("consistency" — largely a tautology).

*Read (1) Edwin Jaynes's *Probability Theory — The Logic of Science* and don't ignore his advice; (2) other pertinent statistics literature.

Reconstruction space (1)

Reconstruction space $\mathcal{R}_0$: A set of ρs such that $p \leftrightarrow \rho$ is a one-to-one mapping.

Example 1: Qubit states $\rho = \frac{1}{2}(1 + x\sigma_x + y\sigma_y + z\sigma_z)$ measured by the 4-outcome crosshair POM with

$$\left.\begin{matrix} p_1 \\ p_2 \end{matrix}\right\} = \frac{1}{4}(1 \pm x)\,, \qquad \left.\begin{matrix} p_3 \\ p_4 \end{matrix}\right\} = \frac{1}{4}(1 \pm y)$$

and constraints $p_1 + p_2 = \frac{1}{2}$, $p_3 + p_4 = \frac{1}{2}$, $p_1^2 + p_2^2 + p_3^2 + p_4^2 \le \frac{3}{8}$.

Example 2: Qubit states measured by the 3-outcome trine POM with

$$p_1 = \frac{1}{3}(1 + x)\,, \qquad \left.\begin{matrix} p_2 \\ p_3 \end{matrix}\right\} = \frac{1}{6}(2 - x \pm \sqrt{3}y)$$

and constraints $p_1 + p_2 + p_3 = 1$, $p_1^2 + p_2^2 + p_3^2 \le \frac{1}{2}$.

For both examples, the self-suggesting $\mathcal{R}_0$ is the equatorial disk of the Bloch ball; the data provide *no* information about z.

Reconstruction space (2)

Example 3: Harmonic oscillator measured by the 2-outcome POM with

$$p_1 = \langle 0|\rho|0\rangle\,, \quad p_2 = 1 - p_1$$

and constraint $p_1 + p_2 = 1$.

Here, a reconstruction space consists of all $\rho = |0\rangle p_1 \langle 0| + p_2\rho'$ where ρ' is *any* state with no ground-state component (ρ' could depend on p_2), and the probability space is that of a tossed coin. The data provide only information about the ground-state probability.

General observations:

– Reconstruction space (usually not unique, often not convex)
$\equiv$ Probability space (unique and convex): We work there!

– Because of the quantum constraints, the probability space is usually smaller than that of the K-sided die:

Quantum State Estimation = Classical state estimation with quantum constraints

Point likelihood, MLE, MLR, SCR

Point likelihood: $L(D|\rho) = p_1^{n_1} p_2^{n_2} \cdots p_K^{n_K}$ = the probability of obtaining data D if ρ is the state.

Maximum-likelihood estimator (MLE) $\widehat{\rho}_{\mathrm{ML}}$: That ρ in $\mathcal{R}_0$ for which the data are more likely than for any other state:

$$\max_{\rho} L(D|\rho) = L(D|\widehat{\rho}_{\mathrm{ML}})\,.$$

How can we equip the MLE with error bars? Our answer: Use optimal regions.

Maximum-likelihood region (MLR) $\widehat{\mathcal{R}}_{\mathrm{ML}}$: That region of estimators for which the data are more likely than for any other region of the same pre-chosen size.

Smallest credible region (SCR) $\widehat{\mathcal{R}}_{\mathrm{SC}}$: The smallest region with the pre-chosen credibility.

Size $\equiv$ Prior content

Scenario 1: You have a pre-existing notion of size for regions in $\mathcal{R}_0$? Fine! Scale all sizes such that $\mathcal{R}_0$ has unit size; then assign the same prior content to regions of the same size.

Scenario 2: You do not have a pre-existing notion of region size? Choose the prior of your liking and measure the size of a region by its prior content.

Either way: **Size of a region $\equiv$ Its prior content**.

Notation: The size of region $\mathcal{R}$ is $S_{\mathcal{R}} = \int_{\mathcal{R}} (\mathrm{d}\rho)$ where $(\mathrm{d}\rho)$ is the prior probability of the infinitesimal space element at state ρ.

Reference: M.J. Evans, I. Guttman, T. Swartz, Can. J. Stat. **34**, 113 (2006).

MLRs and SCRs are BLRs (1)

1 Joint probability that ρ is in $\mathcal{R}$ and data D are obtained:

$$\mathrm{prob}(D \wedge \mathcal{R}) = \int_{\mathcal{R}} (\mathrm{d}\rho)\, L(D|\rho)$$

2 Prior likelihood $L(D)$: $\mathrm{prob}(D \wedge \mathcal{R}_0) = L(D) = \int_{\mathcal{R}_0} (\mathrm{d}\rho)\, L(D|\rho)$

3 Normalization: $\sum_D L(D|\rho) = 1\,, \quad \sum_D L(D) = 1$

4 Two factorizations: $\mathrm{prob}(D \wedge \mathcal{R}) = L(D|\mathcal{R})S_{\mathcal{R}} = C_{\mathcal{R}}(D)L(D)$
with the **region likelihood** $L(D|\mathcal{R})$ and the **credibility** $C_{\mathcal{R}}(D)$.

Both are conditional probabilities: The region likelihood is the probability of obtaining the data D if the state is in the region $\mathcal{R}$; the credibility is the probability that the actual state is in the region $\mathcal{R}$ if the data D have been obtained—the posterior probability of the region.

MLRs and SCRs are BLRs (2)

4 Two factorizations: $\mathrm{prob}(D \wedge \mathcal{R}) = L(D|\mathcal{R})S_{\mathcal{R}} = C_{\mathcal{R}}(D)L(D)$

5 MLR: Maximize the region likelihood for given size,

$$\max_{\mathcal{R}} L(D|\mathcal{R}) = L(D|\widehat{\mathcal{R}}_{\mathrm{ML}}) \quad \text{with } S_{\mathcal{R}} = s$$

6 SCR: Minimize the size for given credibility,

$$\min_{\mathcal{R}} S_{\mathcal{R}} = S_{\widehat{\mathcal{R}}_{\mathrm{sc}}} \quad \text{with } C_{\mathcal{R}}(D) = c$$

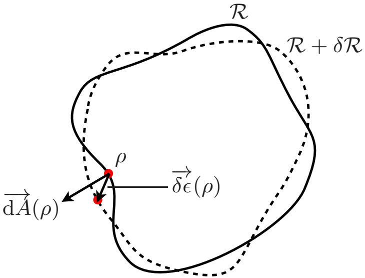
7 These optimization problems are duals of each other:

	MLR	SCR
$S_{\mathcal{R}}$	given	minimize
$\mathrm{prob}(D \wedge \mathcal{R})$	maximize	given

Each MLR is a SCR, each SCR is a MLR.

MLRs and SCRs are BLRs (3)

8 Infinitesimal variation of region $\mathcal{R}$ from a distortion of its boundary $\partial\mathcal{R}$:



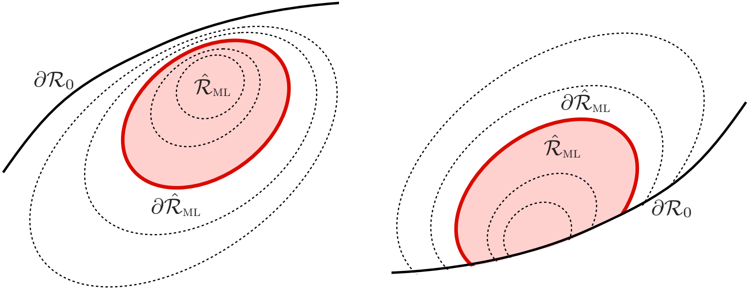
9 Null response of $S_{\mathcal{R}}$ and $\mathrm{prob}(D \wedge \mathcal{R})$:

$$\delta S_{\mathcal{R}} = \int_{\partial\mathcal{R}} \overrightarrow{\mathrm{d}A}(\rho) \cdot \overrightarrow{\delta\epsilon}(\rho) = 0\,,$$

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

MLRs and SCRs are BLRs (4)

10 Requiring that both $\delta S_{\mathcal{R}} = 0$ and $\delta\mathrm{prob}(D \wedge \mathcal{R}) = 0$ implies that the point likelihood $L(D|\rho)$ is constant on $\partial\mathcal{R}$, and larger inside than on the boundary:
The MLRs and the SCRs are **bounded-likelihood regions** (BLRs), which consist of all ρs for which $L(D|\rho)$ exceeds a threshold value.



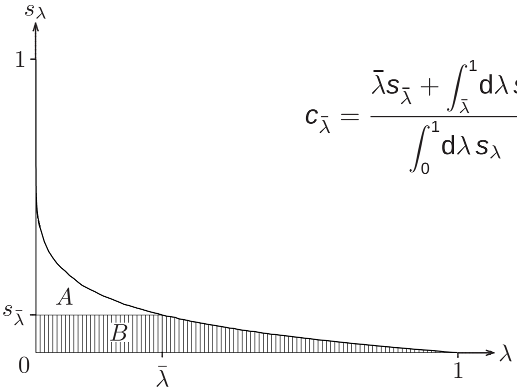
Reference: M.J. Evans, I. Guttman, T. Swartz, Can. J. Stat. **34**, 113 (2006).

MLRs and SCRs are BLRs (5)

11 The set of BLRs is independent of the prior; each BLR contains the MLE.

12 Notation: $\mathcal{R}_\lambda$ is the BLR with $L(D|\rho) \geq \lambda L(D|\widehat{\rho}_{\mathrm{ML}})$; $s_\lambda =$ size of $\mathcal{R}_\lambda$; $c_\lambda =$ credibility of $\mathcal{R}_\lambda$. We have $c_\lambda > s_\lambda$ for $0 < \lambda < 1$.

13 From s_λ to c_λ:



$$c_{\bar\lambda} = \frac{\bar\lambda s_{\bar\lambda} + \int_{\bar\lambda}^{1} \mathrm{d}\lambda\, s_\lambda}{\int_0^1 \mathrm{d}\lambda\, s_\lambda} = \frac{B}{A+B}$$

MLRs and SCRs are BLRs (6)

14 In the limit of $\lambda \to 1$, the BLR $\mathcal{R}_\lambda$ degenerates into the one-point region that contains the MLE, and $c_\lambda \to 0$, $s_\lambda \to 0$, while

$$\frac{c_\lambda}{s_\lambda} \to \frac{L(D|\widehat{\rho}_{\mathrm{ML}})}{L(D)} > 1\,.$$

15 In the limit of $\lambda \to 0$, the $\mathcal{R}_\lambda$ becomes the full reconstruction space $\mathcal{R}_0$, and $c_\lambda \to 1$, $s_\lambda \to 1$.

16 The outcome of the data analysis is reported by communicating the size s_λ and the credibility c_λ as functions of λ. If the probability space is low-dimensional, we can also draw the boundaries of selected BLRs, but this is not possible for high-dimensional $\mathcal{R}_0$s.

Choice of prior

0 Consistency: SCRs should be data-dominated, rather than prior-dominated, for large enough N.

1 Uniformity — a red herring: All priors are uniform.

2 Utility: Be guided by the eventual application.

3 Symmetry: Helpful if used with care.

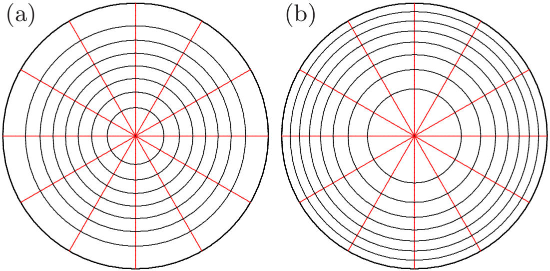
4 Invariance — form invariance, really.

5 Conjugation: Mock posterior for a target state.

6 Marginalization: Convert a prior on the full state space to its marginal on the reconstruction space.

One reference of many: R.E. Kass, L. Wasserman, J. Am. Stat. Assoc. **91**, 1343 (1996)

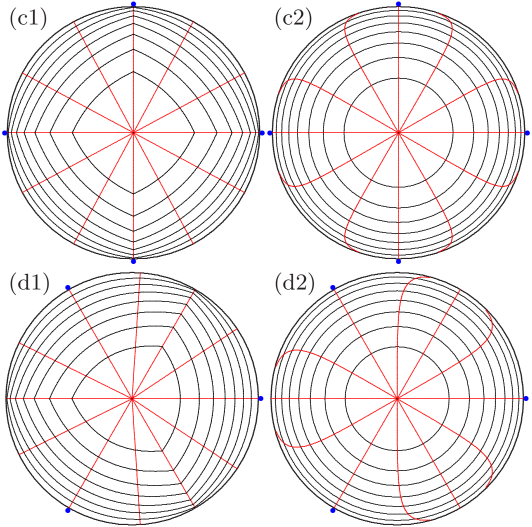
Examples of priors, illustrated by uniform tilings (1)



Tiling (a): A prior in the full-qubit space that is rotationally invariant and uniform in the purity, marginalized onto the unit disk.

Tiling (b): The common primitive qubit prior of the crosshair POM and the trine POM.

Examples of priors, illustrated by uniform tilings (2)

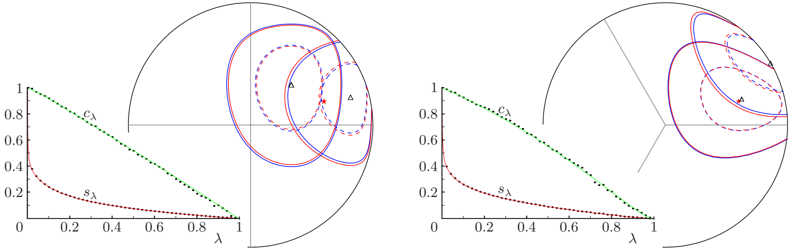


Tilings (c1) and (c2): Jeffreys prior for the crosshair POM.

Tilings (d1) and (d2): Jeffreys prior for the trine POM.

Examples of SCRs: single qubit

$$c_\lambda = \left(\lambda s_\lambda + \int_\lambda^1 \mathrm{d}\lambda'\, s_{\lambda'}\right) \Bigg/ \int_0^1 \mathrm{d}\lambda'\, s_{\lambda'}$$

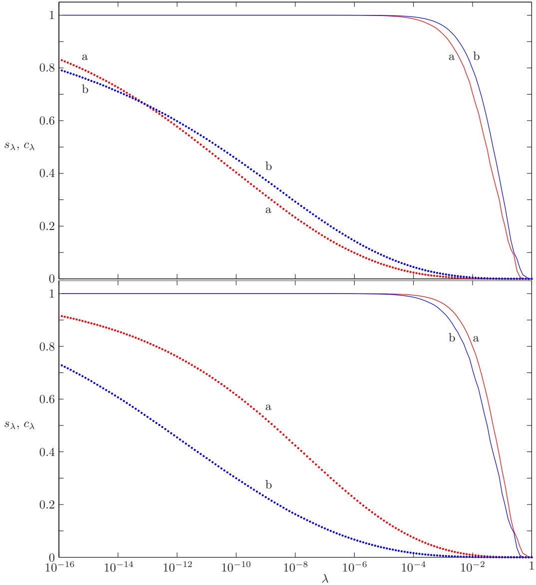


SCRs for credibility $c = 0.5$ and $c = 0.9$; 24 copies measured (in a simulated experiment); primitive (red) and Jeffreys (blue) prior.

(a) crosshair POM: counts $(n_1, n_2, n_3, n_4) = (8, 5, 10, 1)$ and $(6, 3, 10, 5)$

(b) trine POM: counts $(n_1, n_2, n_3) = (15, 8, 1)$ and $(13, 7, 4)$

Examples of SCRs: qubit pairs



Two POMs

top: double crosshair

bottom: trine-antitrine

Two priors

a: primitive prior

b: Jeffreys prior

Note: The data for the top plot are untypical, those for the bottom are typical. The true state is inside the BLR for $\lambda < 3.368 \times 10^{-3}$ (top) and $\lambda < 0.2486$ (bottom).

Take-home messages

1 Quantum state estimation = Classical state estimation with quantum constraints.

Quantum aspects of the problem enter **only** through the Born rule. Except for the implied restrictions on the probabilities, there is no difference between state estimation in quantum mechanics and statistics. Accordingly, **quantum mechanicians can benefit much from methods developed by statisticians.**

2 Bounded-likelihood regions are optimal error regions.

Outlook

1. While we already have efficient methods for calculating the MLE for the data at hand (Many thanks to the Olomouc group!), we still need to further develop the algorithms for computing s_λ, and then c_λ, and thus finding the SCR.
2. For the evaluation of the multi-dimensional integrals for s_λ, one needs good sampling strategies. The bottleneck is the verification that a set of candidate probabilities obeys the quantum constraints. Recent progress: 100 hours of CPU time $\longrightarrow$ 10 hours $\longrightarrow$ 30 minutes.
3. It is possible to reduce the dimensionality of the problem if one is really only interested in a few properties of the state (such as the fidelity with a target state or the concurrence of a two-qubit state).

Discussions with David Nott (Department of Statistics and Applied Probabilty, NUS) are gratefully acknowledged.

THANK YOU