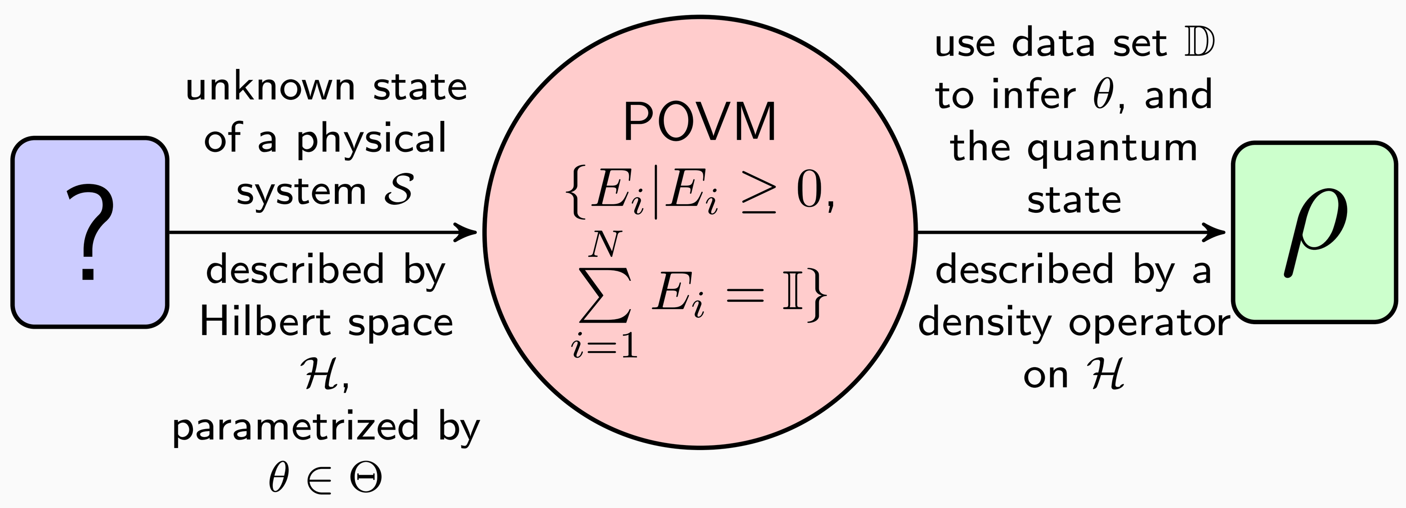
Minimax state estimation under Bregman divergence

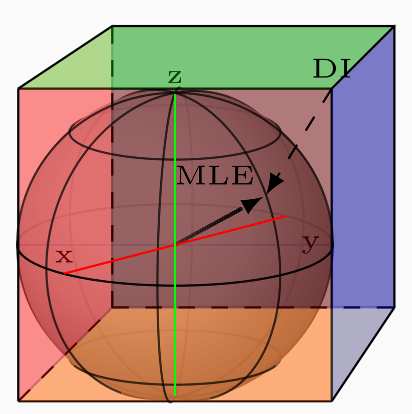
Maria Quadeer, Marco Tomamichel, and Christopher Ferrie (arXiv:1808.08984)

Centre for Quantum Software and Information, University of Technology, Sydney

1. What is quantum state tomography?



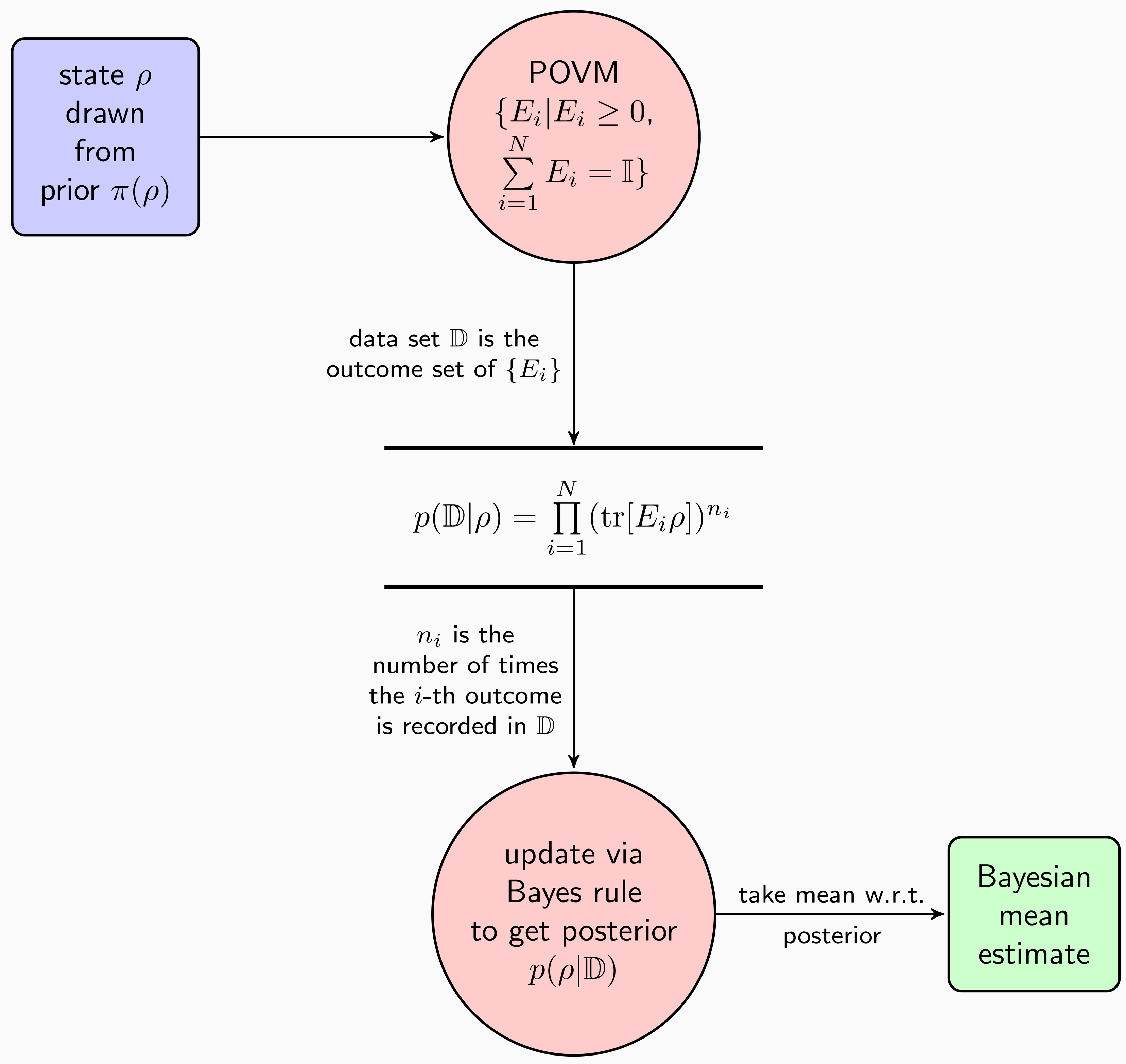
2. Conventional estimation techniques



Direct inversion (DI): Suppose one measures an unknown quantum state in $\mathbb{C}^2$ along the x, y and z directions. Assuming that each of the measurements are performed only once, let us suppose that each of the outcome is 'up' so that $p_x = p_y = p_z = 1$. Now, an estimator that would yield the same probabilities would be the one with the Bloch vector: $(2p_x - 1, 2p_y - 1, 2p_z - 1) = (1,1,1)$. This is an invalid quantum state as it lies outside the Bloch ball (and has negative eigenvalues).

Maximum likelihood estimation (MLE): A likelihood functional $\mathcal{L}[\rho] : \mathcal{S}(\mathcal{H}) \mapsto [0,1]$ is the probability of observing a data set $\mathbb{D}$ given that the system is in the state $\rho \in \mathcal{S}(\mathcal{H})$, $\mathcal{L}[\rho] = p(\mathbb{D}|\rho) = \prod_{i=1}^{N} (\mathrm{tr}[E_i \rho])^{n_i}$ where n_i is the number of times the i-th outcome is recorded in $\mathbb{D}$. $\mathcal{L}[\rho]$ is to be maximized over $\mathcal{S}(\mathcal{H})$ to obtain an estimate. In the above example, $\mathcal{L}[\rho] = (1 + r_x)(1 + r_y)(1 + r_z)/6^3$, to be maximized under $\|\vec{r}\| \leq 1$. This implies that $r_x = r_y = r_z = 1/\sqrt{3}$, which corresponds to an estimator that is a pure state.

3. Bayesian mean estimation



Bayesian mean estimation (BME) allows one to update one's estimate in the light of new data (see Block 6).

4. Estimators, loss functions & risk

- An estimator is defined as the map $\hat{\rho} : \mathcal{X} \mapsto \mathcal{S}(\mathcal{H})$.
- The value of $\hat{\rho}(\mathrm{x})$ is the estimate of ρ_θ when the measurement outcome is X=x.
- To measure how good an estimator is one chooses a distance-measure —called the loss function, denoted as $L(\rho_\theta, \hat{\rho}(x))$.
- We choose Bregman divergence as the loss-function that generalizes both relative entropy and Hilbert-Schmidt distance.
- X is a random variable and to make sense of the loss we average it with respect to the conditional distribution of X—$p(x|\theta)$ to obtain the risk of the estimator:

$$R(\rho_\theta, \hat{\rho}) = \int_{\mathcal{X}} \mathrm{d}p(x|\theta) L(\rho_\theta, \hat{\rho}(x)).$$

5. Bregman divergence

Let $f : [0,1] \mapsto \mathbb{R}$ be a strictly convex continuously-differentiable real-valued function. Then, the Bregman divergence between density matrices ρ, σ is defined as

$$D_f(\rho, \sigma) = \mathrm{tr}\left(f(\rho) - f(\sigma) - f'(\sigma)(\rho - \sigma)\right).$$

- Not a metric but $D_f(\rho, \sigma) \geq 0$ with equality if and only if $\rho = \sigma$.
- Bregman divergence generalizes two important classes of distance-measures.
- Relative entropy obtained by choosing $f : x \mapsto x \log x$.
- Hilbert-Schmidt distance (Schatten 2-norm) obtained by choosing $f : x \mapsto x^2$.

6. Bayes & Minimax estimator

- Risk is a function of ρ_θ.
- There does not exist a unique estimator $\hat{\rho}$ that will minimize risk for all ρ_θ.
- There are two ways of solving this problem—either minimizing average risk or worst-case risk:

$$\underbrace{\text{average risk: } \int_{\Theta} \mathrm{d}\pi(\theta) R(\rho_\theta, \hat{\rho})}_{\text{minimizing estimator: Bayes}} \quad \Bigg| \quad \underbrace{\text{worst-case risk: } \sup_{\theta} R(\rho_\theta, \hat{\rho})}_{\text{minimizing estimator: minimax}}$$

- Bayes estimator is the mean if the loss function is Bregman divergence.

7. How far can Bayesian estimation take us?

Result 1. For any estimator $\hat{\rho}$, there always exists a sequence of Bayes estimators such that the limit of the sequence performs at least as well as $\hat{\rho}$, i.e.

$$R(\rho_\theta, \hat{\rho}) \geq R\left(\rho_\theta, \lim_{n\to\infty} \hat{\rho}_B^{\pi_n}\right), \quad \forall \theta \in \Theta.$$

(Generalization of the work of [1] to Bregman divergence.)

8. Is there a Bayes estimator that is also minimax?

Choice of bad priors can always result in nonsensical estimates, but minimax analysis leads to a natural identification of priors.

Result 2. There always exists a sequence of priors such that the limit of the sequence maximizes the average risk of a Bayes estimator,

$$r(\pi, \hat{\rho}_B^\pi) = \int_{\Theta} \mathrm{d}\pi(\theta) R(\rho_\theta, \hat{\rho}_B^\pi),$$

and the limit of the respective sequence of Bayes estimators minimizes the worst-case risk, i.e., it is minimax. The limit of the sequence of priors is called a least favourable prior. (Generalization of the work of [1] to Bregman divergence.)

9. What is the measurement that minimizes the worst-case risk?

- A POVM that minimizes the worst-case risk is called a minimax POVM.
- In covariant state estimation, given a fixed state ρ_0, one is interested in estimating the states $\rho_\theta \in \{V_g \rho_0 V_g^\dagger\}$, where $g \in G$ is a group element acting on the parameter space Θ, and V_g is the unitary representation G acting on $\mathcal{S}(\mathcal{H})$.

Result 3. Any covariant measurement is minimax for covariant state estimation. Moreover, if there exists a measurement $\mathcal{P}_c$ which is covariant under a subgroup H of G such that $\{V_h | \, h \in H\}$, where V_h is the projective unitary representation of the subgroup H, forms a unitary 2-design, then $\mathcal{P}_c$ is minimax.

- Also, we look at the simplest system of a single qubit (extending the results of [1] to Hilbert-Schmidt distance), and observe that every *spherical 2-design* in $\mathbb{C}^2$ is a minimax POVM.

References

[1] T. Koyama, T. Matsuda, and F. Komaki. "Minimax Estimation of Quantum States Based on the Latent Information Priors". In: *Entropy* 19.11 (2017), p. 618.