Rank–Based Model Selection for Ion Quantum Tomography

Theo Kypraios
http://www.maths.nott.ac.uk/∼tk

School of Mathematical Sciences
University of Nottingham

Mădălin Guță (UoN) and **Ian Dryden** (U South Carolina)

http://arxiv.org/abs/1206.4032

EPSRC

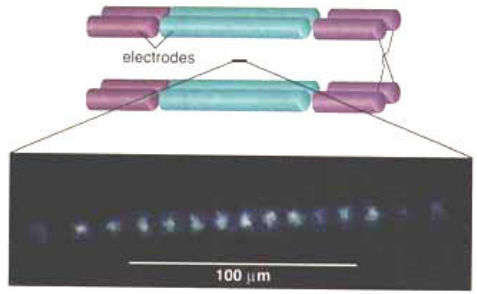
"New Directions in Quantum Statistics"

Nottingham 2012

The University of Nottingham

Quantum Tomography for Trapped Ions

Quantum tomography for trapped ions



[Häffner *et al*, Nature 2005]

Goal: prepare a W (entangled) state of several (4 to 8) ions

Validation: statistical 'reconstruction' of the quantum state $\rho \in M(\mathbb{C}^{2^k})$

- $4^8 - 1 = 65\,535$ parameters to estimate (8 ions)
- $3^8 \times 100 = 656\,100$ repeated measurements
- 10 hours measurement time
- weeks to compute "error bars"
- fidelity between estimator and target state between 0.85 and 0.72

Measurement procedure and statistical model

All measurements are performed on independent identically prepared states $\rho \in M(\mathbb{C}^{2^k})$

1. For each ion choose a spin direction to measure $\sigma_d \in \{\sigma_x, \sigma_y, \sigma_z\}$
2. measure each ion and obtain outcome $\mathbf{s} := (s_1, \ldots, s_k) \in \{1, -1\}^k$

$$\begin{aligned} \mathbb{P}_\rho(\mathbf{s}|\mathbf{d}) &= \mathbb{P}_\rho(s_1, \ldots, s_k | \sigma_{d_1}, \ldots, \sigma_{d_k}) \\ &= \left\langle e_{d_1}^{s_1} \otimes \cdots \otimes e_{d_k}^{s_k} \right| \rho \left| e_{d_1}^{s_1} \otimes \cdots \otimes e_{d_k}^{s_k} \right\rangle \end{aligned}$$

3. Repeat n times and collect counts of outcomes $\{N_{\mathbf{s},\mathbf{d}} : \mathbf{s} \in \{1, -1\}^k\}$

$$\mathbb{P}_\rho\left(\{N_{\mathbf{s},\mathbf{d}} : \mathbf{s} \in \{1, -1\}^k\}\right) = \frac{n!}{\prod_s N_{\mathbf{s},\mathbf{d}}!} \prod_{\mathbf{s}} \mathbb{P}_\rho(\mathbf{s}|\mathbf{d})^{N_{\mathbf{s},\mathbf{d}}}$$

4. Repeat over all 3^k choices of measurement set-ups

Total set of $3^k \times 2^k \gg 4^k$ projections is highly overcomplete in $M(\mathbb{C}^{2^k})$!

Measurement procedure and statistical model

All measurements are performed on independent identically prepared states $\rho \in M(\mathbb{C}^{2^k})$

1. For each ion choose a spin direction to measure $\sigma_d \in \{\sigma_x, \sigma_y, \sigma_z\}$
2. measure each ion and obtain outcome $\mathbf{s} := (s_1, \ldots, s_k) \in \{1, -1\}^k$

$$\begin{aligned} \mathbb{P}_\rho(\mathbf{s}|\mathbf{d}) &= \mathbb{P}_\rho(s_1, \ldots, s_k | \sigma_{d_1}, \ldots, \sigma_{d_k}) \\ &= \left\langle e_{d_1}^{s_1} \otimes \cdots \otimes e_{d_k}^{s_k} \right| \rho \left| e_{d_1}^{s_1} \otimes \cdots \otimes e_{d_k}^{s_k} \right\rangle \end{aligned}$$

3. Repeat n times and collect counts of outcomes $\{N_{\mathbf{s},\mathbf{d}} : \mathbf{s} \in \{1, -1\}^k\}$

$$\mathbb{P}_\rho\left(\{N_{\mathbf{s},\mathbf{d}} : \mathbf{s} \in \{1, -1\}^k\}\right) = \frac{n!}{\prod_s N_{\mathbf{s},\mathbf{d}}!} \prod_{\mathbf{s}} \mathbb{P}_\rho(\mathbf{s}|\mathbf{d})^{N_{\mathbf{s},\mathbf{d}}}$$

4. Repeat over all 3^k choices of measurement set-ups

Total set of $3^k \times 2^k \gg 4^k$ projections is highly overcomplete in $M(\mathbb{C}^{2^k})$!

Measurement procedure and statistical model

All measurements are performed on independent identically prepared states $\rho \in M(\mathbb{C}^{2^k})$

1. For each ion choose a spin direction to measure $\sigma_d \in \{\sigma_x, \sigma_y, \sigma_z\}$
2. measure each ion and obtain outcome $\mathbf{s} := (s_1, \ldots, s_k) \in \{1, -1\}^k$

$$\begin{aligned} \mathbb{P}_\rho(\mathbf{s}|\mathbf{d}) &= \mathbb{P}_\rho(s_1, \ldots, s_k | \sigma_{d_1}, \ldots, \sigma_{d_k}) \\ &= \left\langle e_{d_1}^{s_1} \otimes \cdots \otimes e_{d_k}^{s_k} \right| \rho \left| e_{d_1}^{s_1} \otimes \cdots \otimes e_{d_k}^{s_k} \right\rangle \end{aligned}$$

3. Repeat n times and collect counts of outcomes $\{N_{\mathbf{s},\mathbf{d}} : \mathbf{s} \in \{1, -1\}^k\}$

$$\mathbb{P}_\rho\left(\{N_{\mathbf{s},\mathbf{d}} : \mathbf{s} \in \{1, -1\}^k\}\right) = \frac{n!}{\prod_s N_{\mathbf{s},\mathbf{d}}!} \prod_{\mathbf{s}} \mathbb{P}_\rho(\mathbf{s}|\mathbf{d})^{N_{\mathbf{s},\mathbf{d}}}$$

4. Repeat over all 3^k choices of measurement set-ups

Total set of $3^k \times 2^k \gg 4^k$ projections is highly overcomplete in $M(\mathbb{C}^{2^k})$!

Measurement procedure and statistical model

All measurements are performed on independent identically prepared states $\rho \in M(\mathbb{C}^{2^k})$

1. For each ion choose a spin direction to measure $\sigma_d \in \{\sigma_x, \sigma_y, \sigma_z\}$
2. measure each ion and obtain outcome $\mathbf{s} := (s_1, \ldots, s_k) \in \{1, -1\}^k$

$$\begin{aligned} \mathbb{P}_\rho(\mathbf{s}|\mathbf{d}) &= \mathbb{P}_\rho(s_1, \ldots, s_k | \sigma_{d_1}, \ldots, \sigma_{d_k}) \\ &= \left\langle e_{d_1}^{s_1} \otimes \cdots \otimes e_{d_k}^{s_k} \right| \rho \left| e_{d_1}^{s_1} \otimes \cdots \otimes e_{d_k}^{s_k} \right\rangle \end{aligned}$$

3. Repeat n times and collect counts of outcomes $\{N_{\mathbf{s},\mathbf{d}} : \mathbf{s} \in \{1, -1\}^k\}$

$$\mathbb{P}_\rho\left(\{N_{\mathbf{s},\mathbf{d}} : \mathbf{s} \in \{1, -1\}^k\}\right) = \frac{n!}{\prod_s N_{\mathbf{s},\mathbf{d}}!} \prod_{\mathbf{s}} \mathbb{P}_\rho(\mathbf{s}|\mathbf{d})^{N_{\mathbf{s},\mathbf{d}}}$$

4. Repeat over all 3^k choices of measurement set-ups

Total set of $3^k \times 2^k \gg 4^k$ projections is highly overcomplete in $M(\mathbb{C}^{2^k})$!

Typical Measurement data

- 3^k columns of length 2^k
- one column for each measurement setting
- each column contains the counts totalling 100, of the 2^k possible outcomes
- frequencies of outcomes are bad estimates of probabilities, but overall there is quite a lot of information.

1	2	11	11	11	21	5	16	21	19	11	16	2	26	15	5
2	19	10	6	15	4	22	10	3	12	8	16	18	5	14	16
3	30	12	15	9	10	18	14	3	6	11	4	4	2	1	5
4	0	4	15	10	17	2	4	14	13	0	4	8	5	1	3
5	21	13	12	7	6	5	14	12	8	12	7	19	3	8	3
6	1	12	14	0	1	1	0	6	6	12	8	2	6	2	7
7	1	2	0	19	7	12	14	6	7	14	7	9	23	15	34
8	0	1	1	0	4	8	0	6	6	0	7	12	4	15	5
9	21	17	8	10	7	7	14	9	8	15	6	9	6	3	0
10	2	16	15	0	12	9	0	3	4	1	7	3	0	4	6
11	0	0	1	17	9	2	14	12	7	0	1	0	5	5	2
12	1	1	1	0	2	8	0	4	3	0	1	0	0	3	1
13	1	0	1	1	0	0	0	0	0	14	9	7	6	2	4
14	0	1	0	0	0	1	0	0	1	1	5	6	0	2	2
15	1	0	0	1	0	0	0	0	0	1	2	0	9	6	3
16	0	0	0	0	0	0	0	1	0	0	0	1	0	4	4

[Data set 4 ions (from T. Monz)]

Motivation

- Why does it work?
- What is the structure of the data? Are we in an asymptotic regime?
- Are there other "less expensive" estimation methods?
- What are the "error bars" (aka quantify parameter uncertainty)?

In addition, we are

- Interested in alternative methods aimed at reducing the dimensionality of the state estimation problem ...
- ... without making unwarranted and unrealistic assumptions.

Motivation

- Why does it work?
- What is the structure of the data? Are we in an asymptotic regime?
- Are there other "less expensive" estimation methods?
- What are the "error bars" (aka quantify parameter uncertainty)?

In addition, we are

- Interested in alternative methods aimed at reducing the dimensionality of the state estimation problem ...
- ... without making unwarranted and unrealistic assumptions.

Motivation

- Why does it work?
- What is the structure of the data? Are we in an asymptotic regime?
- Are there other "less expensive" estimation methods?
- What are the "error bars" (aka quantify parameter uncertainty)?

In addition, we are

- Interested in alternative methods aimed at reducing the dimensionality of the state estimation problem ...
- ... without making unwarranted and unrealistic assumptions.

Motivation

- Why does it work?
- What is the structure of the data? Are we in an asymptotic regime?
- Are there other "less expensive" estimation methods?
- What are the "error bars" (aka quantify parameter uncertainty)?

In addition, we are

- Interested in alternative methods aimed at reducing the dimensionality of the state estimation problem ...
- ... without making unwarranted and unrealistic assumptions.

Motivation

- Why does it work?
- What is the structure of the data? Are we in an asymptotic regime?
- Are there other "less expensive" estimation methods?
- What are the "error bars" (aka quantify parameter uncertainty)?

In addition, we are

- Interested in alternative methods aimed at reducing the dimensionality of the state estimation problem ...
- ... without making unwarranted and unrealistic assumptions.

Motivation

- Why does it work?
- What is the structure of the data? Are we in an asymptotic regime?
- Are there other "less expensive" estimation methods?
- What are the "error bars" (aka quantify parameter uncertainty)?

In addition, we are

- Interested in alternative methods aimed at reducing the dimensionality of the state estimation problem ...
- ... without making unwarranted and unrealistic assumptions.

Statistical Inferece (aka State Estimation)

Fisher information, Cramér-Rao bound and asymptotic normality

Let $X_1, \ldots, X_n$ be i.i.d. data with probability distribution $\mathbb{P}_\theta$ and $\theta \in \mathbb{R}^p$

- Fisher information matrix $I(\theta)$

$$I(\theta)_{i,j} := \int \frac{\partial p_\theta(x)}{\partial \theta_i} \frac{\partial p_\theta(x)}{\partial \theta_i} p_\theta(x) dx$$

- Cramér-Rao bound: for any unbiased estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$

$$\mathrm{Var}(\hat{\theta}_n) \geq I(\theta)^{-1}$$

- Efficient estimators are asymptotically normal (e.g. MLE)

$$\sqrt{n}(\hat{\theta}_n - \theta) \approx N(0, I^{-1}(\theta))$$

- Risk for quadratic loss function

$$n\mathbb{E}\left[(\hat{\theta}_n - \theta)^t G(\hat{\theta}_n - \theta)\right] \to \mathrm{Tr}(GI(\theta)^{-1})$$

Fisher information, Cramér-Rao bound and asymptotic normality

Let $X_1, \ldots, X_n$ be i.i.d. data with probability distribution $\mathbb{P}_\theta$ and $\theta \in \mathbb{R}^p$

- Fisher information matrix $I(\theta)$

$$I(\theta)_{i,j} := \int \frac{\partial p_\theta(x)}{\partial \theta_i} \frac{\partial p_\theta(x)}{\partial \theta_i} p_\theta(x) dx$$

- Cramér-Rao bound: for any unbiased estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$

$$\mathrm{Var}(\hat{\theta}_n) \geq I(\theta)^{-1}$$

- Efficient estimators are asymptotically normal (e.g. MLE)

$$\sqrt{n}(\hat{\theta}_n - \theta) \approx N(0, I^{-1}(\theta))$$

- Risk for quadratic loss function

$$n\mathbb{E}\left[(\hat{\theta}_n - \theta)^t G(\hat{\theta}_n - \theta)\right] \to \mathrm{Tr}(GI(\theta)^{-1})$$

Estimation of Pure States

- Pure states are arguably the golden standard of the most state preparation system.
- Do a simulation study where:
  - the state is assumed to be pure;
  - estimate it using Multiple Ions Tomography.
- Why?
  - get more insight into the statistical power of the measurement set-up;
  - investigate the asymptotic properties;
  - and prepare the ground for the results to follow when the state is fitted to models of increasing rank.

Estimation of Pure States

- Pure states are arguably the golden standard of the most state preparation system.
- Do a simulation study where:
  - the state is assumed to be pure;
  - estimate it using Multiple Ions Tomography.
- Why?
  - get more insight into the statistical power of the measurement set-up;
  - investigate the asymptotic properties;
  - and prepare the ground for the results to follow when the state is fitted to models of increasing rank.

Estimation of Pure States

- Pure states are arguably the golden standard of the most state preparation system.
- Do a simulation study where:
  - the state is assumed to be pure;
  - estimate it using Multiple Ions Tomography.
- Why?
  - get more insight into the statistical power of the measurement set-up;
  - investigate the asymptotic properties;
  - and prepare the ground for the results to follow when the state is fitted to models of increasing rank.

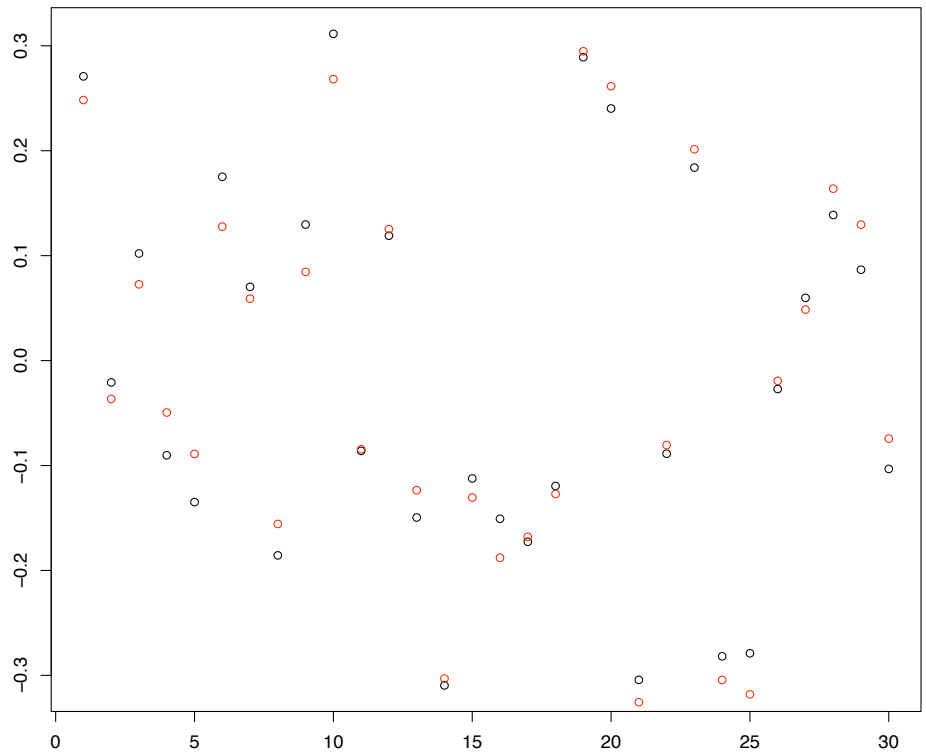
Estimation of Pure States

- Pure states are arguably the golden standard of the most state preparation system.
- Do a simulation study where:
  - the state is assumed to be pure;
  - estimate it using Multiple Ions Tomography.
- Why?
  - get more insight into the statistical power of the measurement set-up;
  - investigate the asymptotic properties;
  - and prepare the ground for the results to follow when the state is fitted to models of increasing rank.

Maximum Likelihood Estimation (Rank 1)

Maximum likelihood estimator

$$\hat{\rho} := \arg\max_{\tau \in \mathcal{S}_k} \sum_{\mathbf{s},\mathbf{d}} N(\mathbf{s}|\mathbf{d}) \log \mathbb{P}_\tau(\mathbf{s}|\mathbf{d}), \tag{1}$$



Parameters of true (black) versus estimated (red) pure state of 4 ions

Maximum Likelihood Estimation (Rank 1)

As a figure of merit we consider the mean square error (MSE)

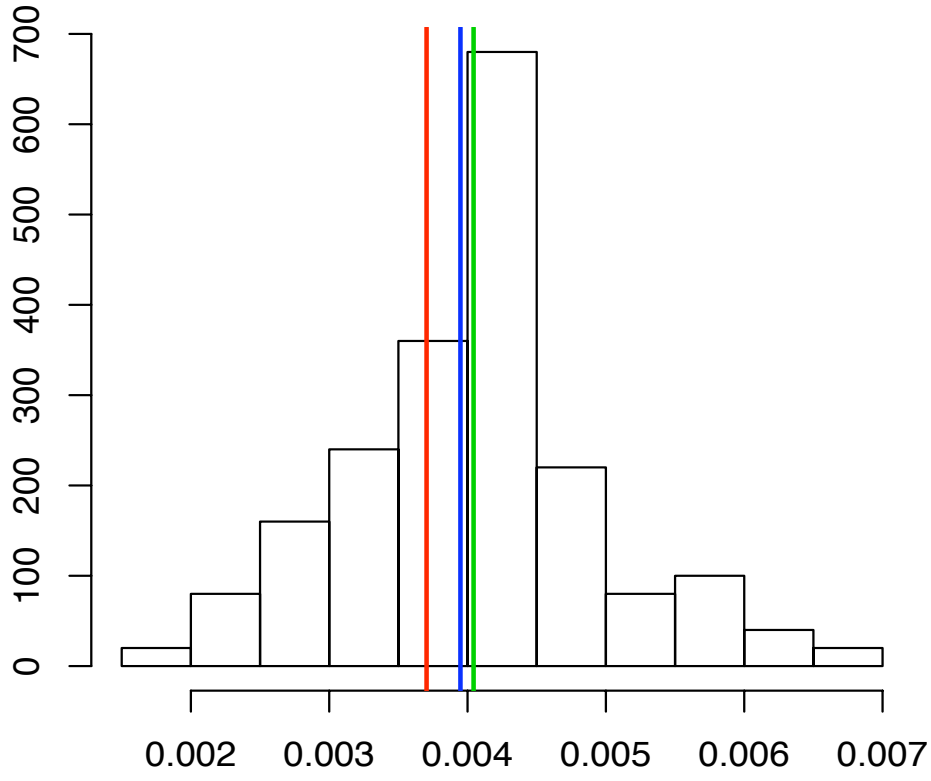
$$MSE(\hat{\rho}) := \mathbb{E}(\||\rho - \hat{\rho}\|_2^2)$$

with the norm-two distance squared defined as

$$\|\rho - \hat{\rho}\|_2^2 := \sum_{i,j=1}^{2^k} |\rho_{i,j} - \hat{\rho}_{i,j}|^2 = \sum_{\mathbf{i}} |\rho_{\mathbf{i}} - \hat{\rho}_{\mathbf{i}}|^2.$$

where $\rho_{\mathbf{i}}$ are the Fourier coefficients.

Maximum Likelihood Estimation (Rank 1)



Histogram of $\|\hat{\rho}_{ml} - \rho\|_2^2$ for a pure state $\rho \in M(\mathbb{C}^{2^4})$ (100 repetitions)

- Very good agreement with asymptotic theory
  - mean (green line) very close to the Cramer-Rao bound (blue line)
- Measurement is very close to optimal
  - Quantum CR bound $30/(100 * 3^4) = 0.0037$ (red line) slightly smaller than classical CR

MLE (Rank 1): Summary

- The MSE of the MLE shows good agreement with asymptotic theory for $(n = 100, k = 4)$;
- Even for larger k and fixed n we expect similarly good agreement due to the relationship of the number of samples $(n \cdot 3^k)$ with respect to the number of parameters $2(2^k - 1)$.
- The multiple ions measurement set-up appears to be quasi-optimal; adaptive strategies for choosing the settings cannot offer a significant improvement.

Full Data vs Coarse Grained Data

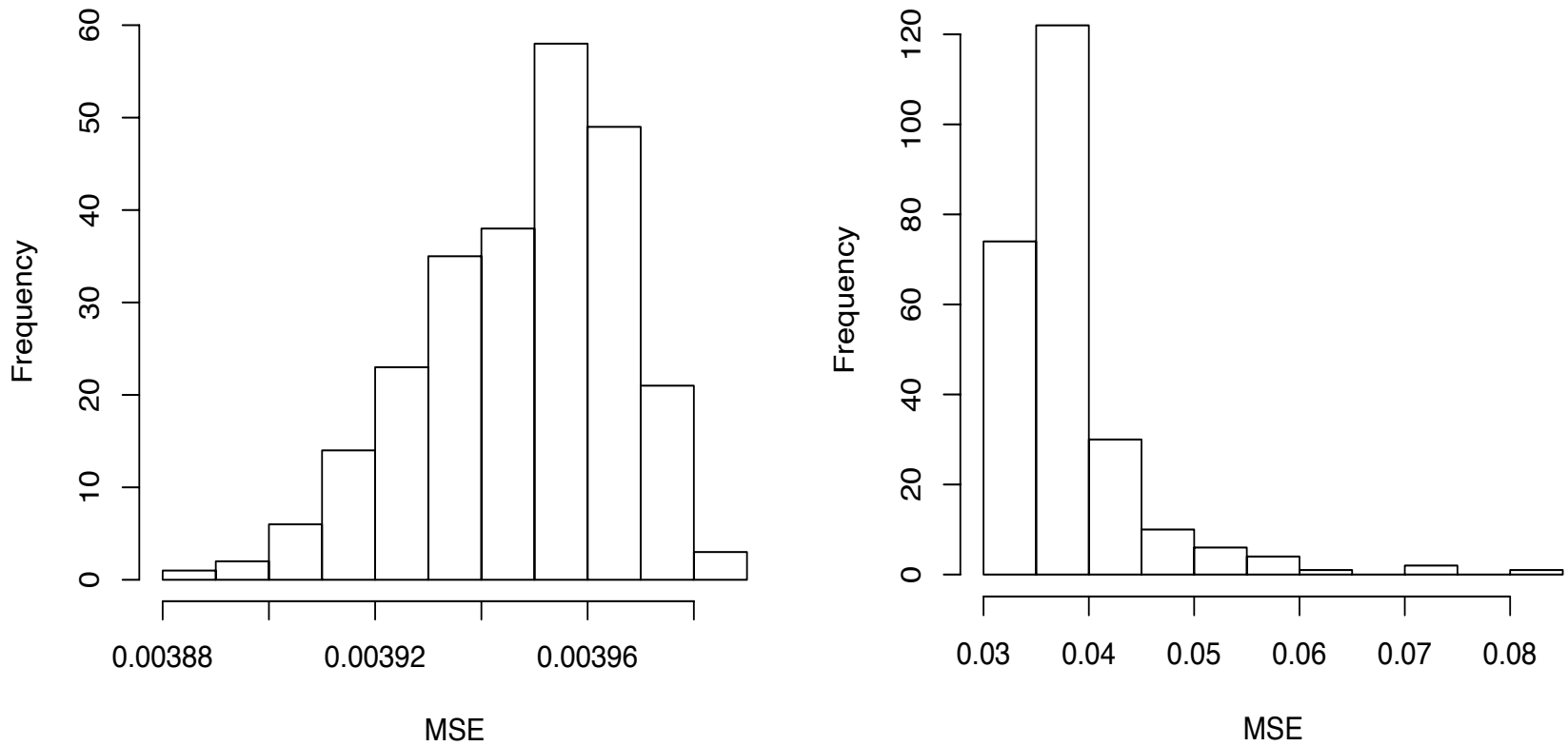


Figure: Histograms of asymptotic MSE's for 250 randomly chosen pure states, with 4 ions and 100 measurements. Left panel: full counts dataset. Right panel: coarse grained dataset. Keeping only the empirical means of the Pauli products leads to a 10 fold increase in the MSE.

Model Selection

Model Selection for Quantum Tomography

- So far we have looked at the extreme scenarios of "full" quantum tomography and estimation of pure states.
- In reality, the states produced in experiments tend to have one or few significant eigenvalues and a large number of small eigenvalues of different orders of magnitude, which account for the imperfections in the preparation procedure.

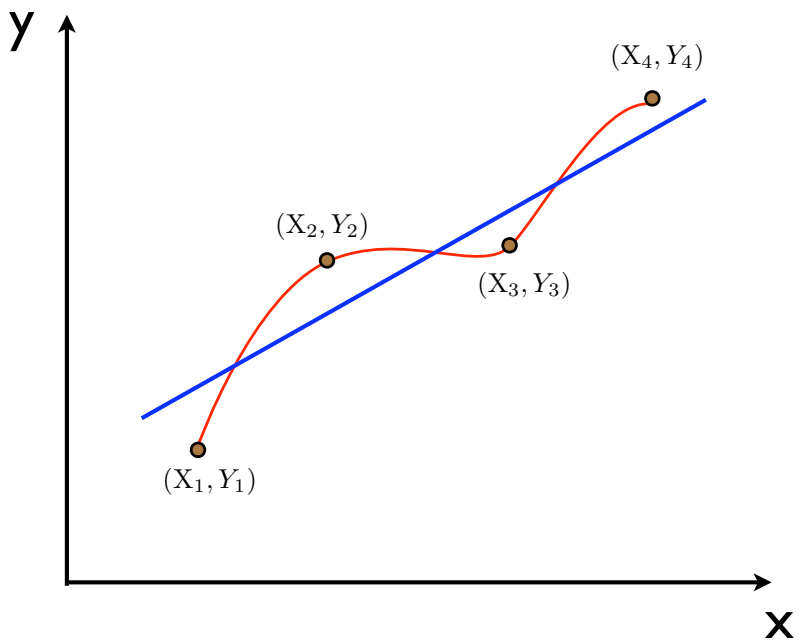
Therefore, neither of the two settings seems to be suitable:

- the former underfits the real state while
- the latter overfits by trying to estimate eigenvalues that may not be statistically significant.

Model selection in Statistics

- Many statistical problems deal with model selection, i.e. how to choose the 'right' description of observed data from a family of models of increasing complexity
- Non-linear regression: want to identify $f : \mathbb{R} \to \mathbb{R}$ from noisy observations (X_i, Y_i)

$$Y_i = f(X_i) + \epsilon_i, \qquad i = 1, \ldots, n$$



Underfitting versus overfitting?

Model selection in Statistics

- Avoid over-fitting using penalised maximum likelihood estimation . . .
- . . . where the penalty measures the complexity of the model (e.g. the number of non-zero coefficients with respect to an appropriate basis).
- Design a class of statistical models with various degrees of complexity, and decide which model is most suitable for describing the data.
- **Aim**: apply model selection methodology to state tomography, the models being the families of states of a given rank.
    - Akaike Information Criterion (AIC)
    - Bayesian Information Criterion (BIC).

Selecting the Rank of the State Estimator

If the state is **not** known to be pure, can we estimate it without doing ML over **all** states?

- in many experiments the engineered states have a good degree of purity;
- often one prefers a parsimonious explanation of the observations (Occam's razor)
- computing full rank estimator may be more challenging.

Model selection: Based on the data $\{N_{\mathbf{s},\mathbf{d}}\}$ choose a model $\mathcal{S}(\hat{r})$ and an estimator $\hat{\rho} \in \mathcal{S}(\hat{r})$

- how to do this consistently ($\hat{r} \approx \mathrm{rank}(\rho)$)?
- how to do this accurately ($\|\rho - \hat{\rho}\|_2^2$ is small)?

AIC and BIC for Model Selection

The **AIC** for model $\mathcal{M}_r$ is

$$\text{AIC}(r) = -2 \cdot \ell_{\widehat{\theta}_r} + 2p(r),$$

and the chosen model is the one with the minimum AIC.

- Since $p(r)$ is larger for more complex models, the AIC formally biases against overly complicated models.

The **BIC** for model $\mathcal{M}_r$ is defined as

$$\text{BIC}(r) = -2 \cdot \ell_{\hat{\theta}_r} + p(r) \log(n)$$

where n is the sample size and the chosen model is the one with the minimum BIC.

- BIC differs from the AIC only in the second term which increases with n, so that BIC favors simpler models (that is models with a smaller number of parameters) compared to AIC.

Choosing the rank of the state by BIC /AIC

1. Perform separate ML over states of rank $r = 1, 2, 3, \ldots$ to obtain $\hat{\rho}_{\mathrm{ml}}^{(1)}, \hat{\rho}_{\mathrm{ml}}^{(2)}, \hat{\rho}_{\mathrm{ml}}^{(3)} \ldots$

2. Choose the rank r for which $\hat{\rho}_{\mathrm{ml}}^{(r)}$ achieves the 'best' likelihood-complexity trade-off:

$$\text{TRADE-OFF}(r) := -2 \log \mathbb{P}_{\hat{\rho}_{\mathrm{ml}}^{(r)}}(\text{DATA}) + \text{PENALTY}(r, N)$$

- Akaike information criterion (AIC)

$$\text{PENALTY}(r, n) := 2 * \sharp \text{ parameters}(\hat{\rho}_{\mathrm{ml}}^{(r)}) = 2(2r * 2^k - r^2 - 1)$$

- Bayesian information criterion (BIC)

$$\text{PENALTY}(r, N) := \sharp \text{ parameters}(\hat{\rho}_{\mathrm{ml}}^{(r)}) * \ln(N) = (2r * 2^k - r^2 - 1) * \ln(N)$$

Choosing the rank of the state by BIC /AIC

1. Perform separate ML over states of rank $r = 1, 2, 3, \ldots$ to obtain $\hat{\rho}^{(1)}_{\mathrm{ml}}, \hat{\rho}^{(2)}_{\mathrm{ml}}, \hat{\rho}^{(3)}_{\mathrm{ml}} \ldots$

2. Choose the rank r for which $\hat{\rho}^{(r)}_{\mathrm{ml}}$ achieves the 'best' likelihood-complexity trade-off:

$$\text{TRADE-OFF}(r) := -2 \log \mathbb{P}_{\hat{\rho}^{(r)}_{\mathrm{ml}}}(\text{DATA}) + \text{PENALTY}(r, N)$$

- Akaike information criterion (AIC)

$$\text{PENALTY}(r, n) := 2 * \sharp \text{ parameters}(\hat{\rho}^{(r)}_{\mathrm{ml}}) = 2(2r * 2^k - r^2 - 1)$$

- Bayesian information criterion (BIC)

$$\text{PENALTY}(r, N) := \sharp \text{ parameters}(\hat{\rho}^{(r)}_{\mathrm{ml}}) * \ln(N) = (2r * 2^k - r^2 - 1) * \ln(N)$$

Choosing the rank of the state by BIC /AIC

1. Perform separate ML over states of rank $r = 1, 2, 3, \ldots$ to obtain $\hat{\rho}^{(1)}_{\mathrm{ml}}, \hat{\rho}^{(2)}_{\mathrm{ml}}, \hat{\rho}^{(3)}_{\mathrm{ml}} \ldots$

2. Choose the rank r for which $\hat{\rho}^{(r)}_{\mathrm{ml}}$ achieves the 'best' likelihood-complexity trade-off:

$$\mathrm{TRADE\text{-}OFF}(r) := -2 \log \mathbb{P}_{\hat{\rho}^{(r)}_{\mathrm{ml}}}(\mathrm{DATA}) + \mathrm{PENALTY}(r, N)$$

- Akaike information criterion (AIC)

$$\mathrm{PENALTY}(r, n) := 2 * \sharp \text{ parameters}(\hat{\rho}^{(r)}_{\mathrm{ml}}) = 2(2r * 2^k - r^2 - 1)$$

- Bayesian information criterion (BIC)

$$\mathrm{PENALTY}(r, N) := \sharp \text{ parameters}(\hat{\rho}^{(r)}_{\mathrm{ml}}) * \ln(N) = (2r * 2^k - r^2 - 1) * \ln(N)$$

Simulation Studies

Study 1: Randomly Chosen Low Rank States

- 3 "random" states of ranks 1, 2, and 3 of $k = 4$ ions
- generated 100 datasets from each state;
- each dataset with $n = 100$ measurement repetitions.

We then computed the maximum likelihood estimators (MLEs) for the ranks between 1 and 4 and used AIC and BIC to select the optimal rank.

	AIC rank			
true rank	1	2	3	4
1	82	9	9	0
2	0	74	26	0
3	0	1	80	19

	BIC rank			
true rank	1	2	3	4
1	99	0	1	0
2	7	90	3	0
3	0	5	95	0

Summary: BIC selected the correct rank for each state in roughly 90% of the cases while for AIC the rate is about 80%

While at first sight this may appear to be a surprisingly good performance, we can show that it agrees very well with the predictions of asymptotic theory.

Study 2: One Ion simulations

We have seen that the performance of the model selection criteria depends on the spectrum of eigenvalues of the true state, and on the number of measurement repetitions.

To investigate this dependence we performed a statistical experiment with three one-ion states ($k = 1$) of different degrees of purity:

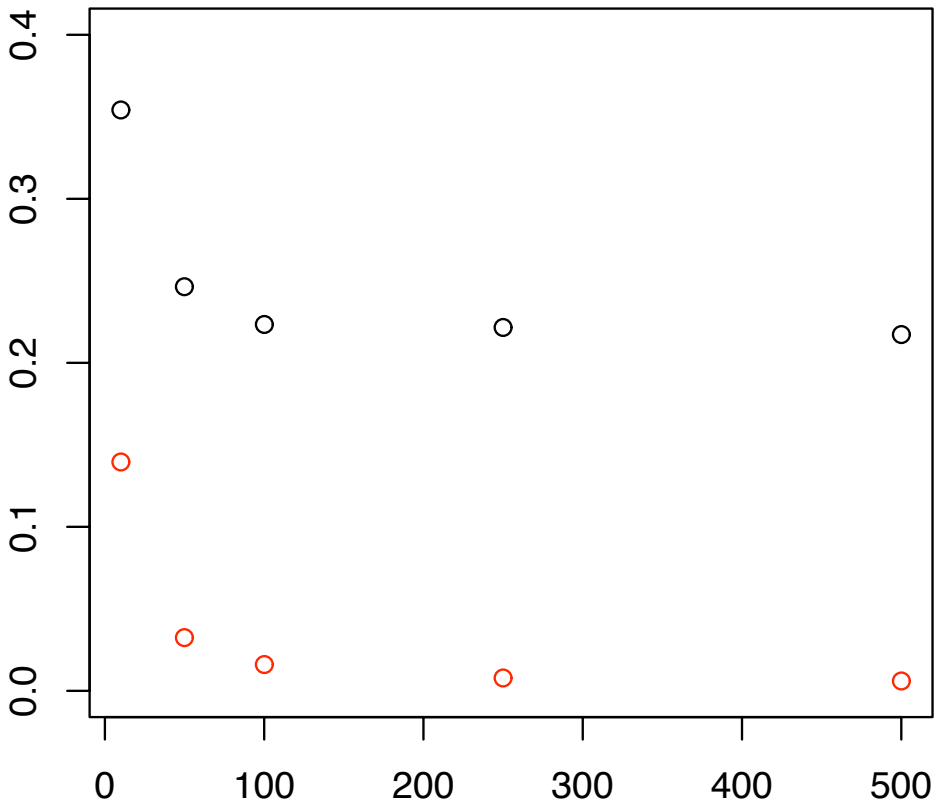
- a pure state;
- one with eigenvalues $(0.95, 0.05)$;
- and the other with eigenvalues $(0.72, 0.28)$.

For each state we simulated datasets with varying number of repetitions $n = 10, 50, 100, 250, 500$.

We count the number of times (out of 1000 samples) BIC and AIC choose the correct rank of the state, for all possible choices of states and measurement repetitions.

Study 2: Mixed state

- True state has eigenvalues $(0.72, 0.28)$
- Both BIC and AIC guess the correct rank 2 from a small number of measurements



Mean square error for rank 1 and 2 estimators

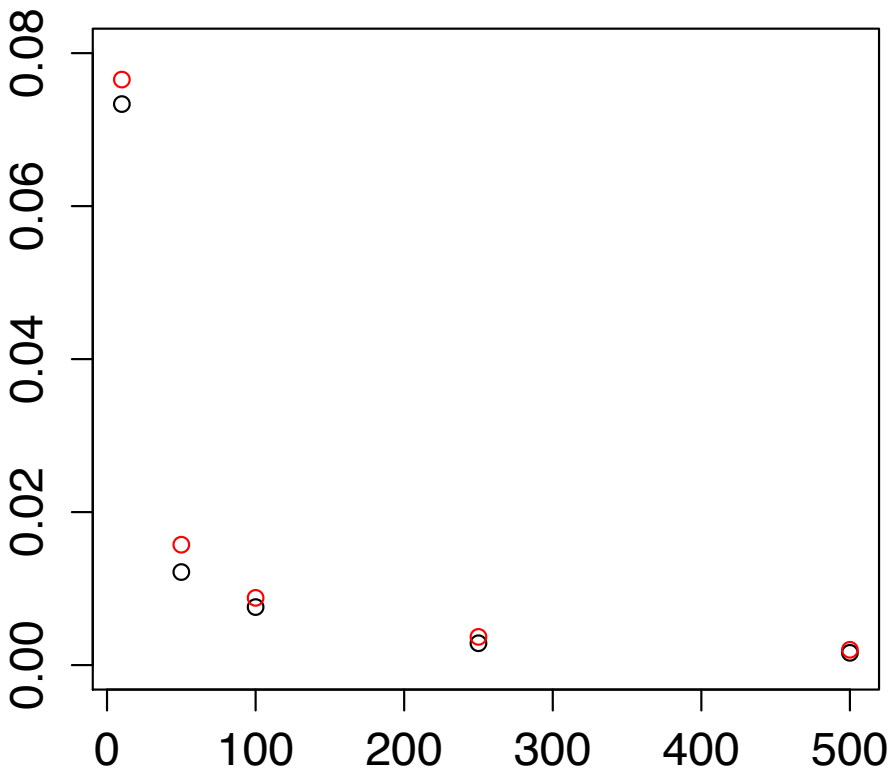
Measurement repetitions

	10	50	100	250	500
BIC	384	973	998	997	988
AIC	594	992	998	997	988

Counts of correct (rank two) estimators for BIC and AIC out of 1000

Study 2: Pure State

- True state is pure
- Both BIC and AIC guess the correct rank for all n
- BIC is slightly more consistent than AIC



Mean square error for rank 1 and 2 estimators

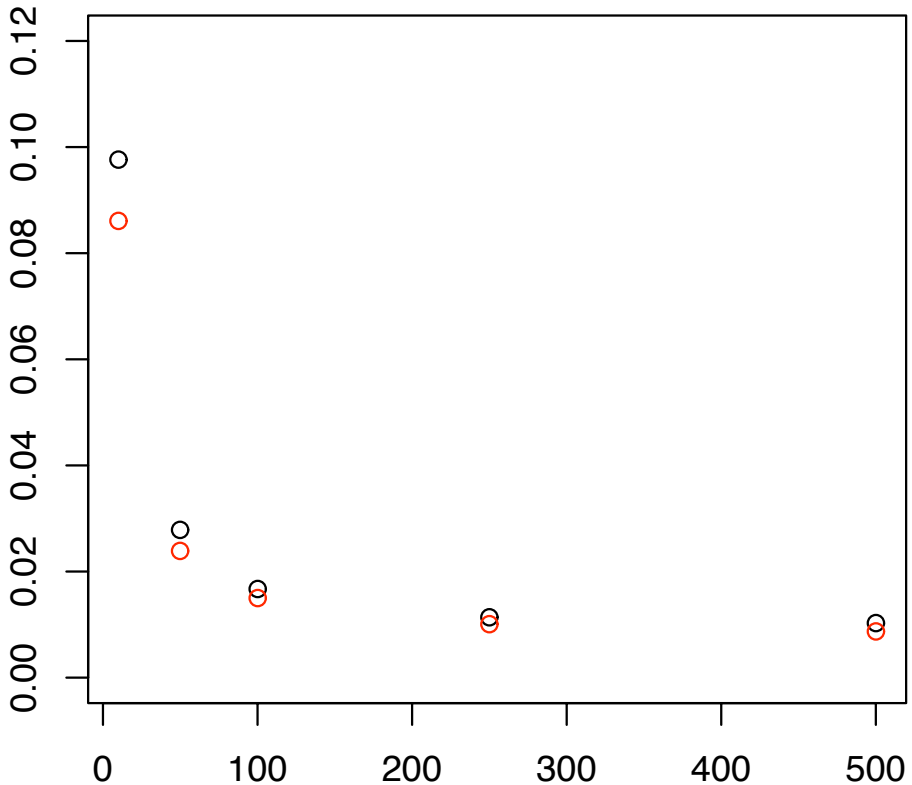
Measurement repetitions

	10	50	100	250	500
BIC	987	990	994	992	996
AIC	945	944	919	927	930

Counts of correct (rank one) estimators for BIC and AIC out of 1000

Study 2: Almost Pure State

- True state has eigenvalues $(0.95, 0.05)$
- BIC is slower than AIC in guessing the correct rank 2



Mean square error for rank 1 and 2 estimators

Measurement repetitions

	10	50	100	250	500
BIC	25	83	183	394	706
AIC	77	312	502	802	942

Counts of correct (rank two) estimators for BIC and AIC out of 1000

Application to (Real) Experimental Data

- Measurements consisted of counts for the 3^4 measurement settings, with a number $n = 4800$ of repetitions for each setting.
- the aim of the experiment was to create a particular 4 ions bound entangled state of rank 4 ("Smolin state").

Model Selection for 4 ions Experimental Data

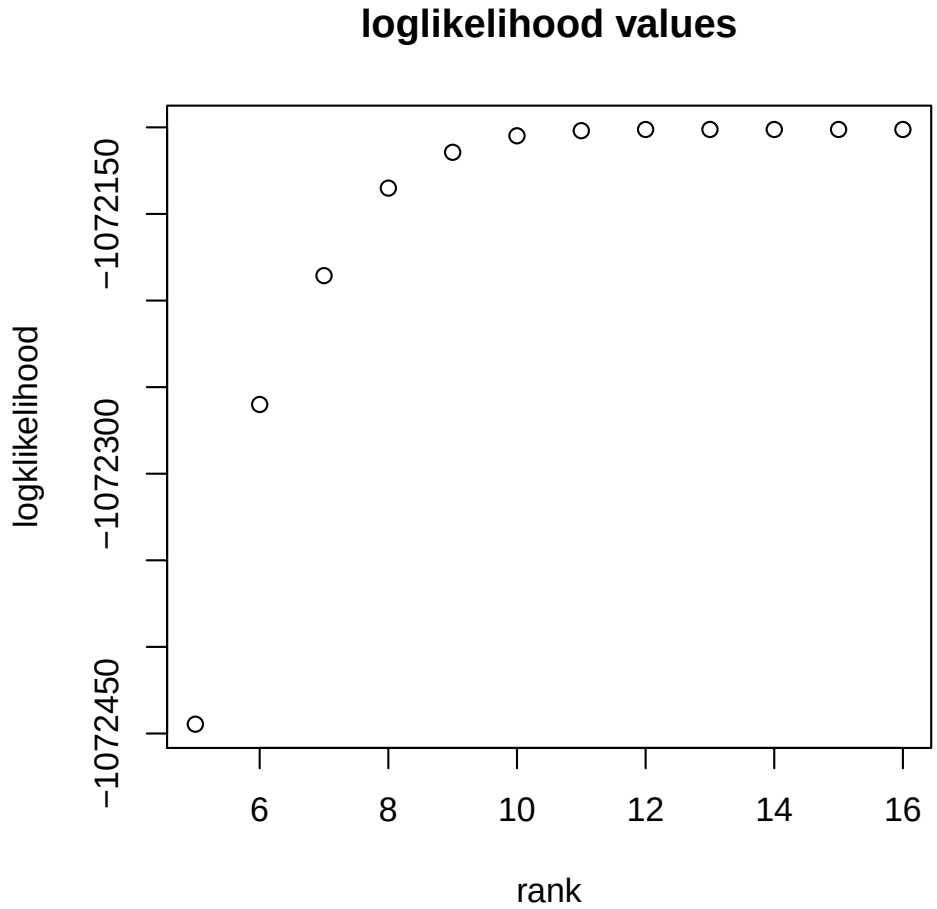


Figure: Log-likelihood values for the maximum likelihood estimator as a function of rank.

Model Selection for 4 ions Experimental Data

RANK	AIC	BIC	EIGENVALUES MLE RANK 10	EIGENVALUES MLE RANK 16
1	2397395	2397722	2.337 e-01	2.332 e-01
2	2217096	2217738	2.290 e-01	2.277 e-01
3	2170638	2171573	2.258 e-01	2.253 e-01
4	2146295	2147502	1.725 e-01	1.721 e-01
5	2145157	2146614	4.599 e-02	4.487 e-02
6	2144830	**2146515**	2.656 e-02	2.445 e-02
7	2144719	2146611	2.385 e-02	2.229 e-02
8	2144652	2146728	1.948 e-02	1.884 e-02
9	**2144641**	2146880	1.226 e-02	1.155 e-02
10	2144648	2147028	1.067 e-02	1.001 e-02
11	2144664	2147164	0	6.057 e-03
12	2144680	2147279	0	2.751 e-03
13	2144694	2147369	0	6.779 e-04
14	2144704	2147433	0	5.278 e-06
15	2144710	2147472	0	2.153 e-06
16	2144712	2147484	0	1.702 e-16

Is the State of Rank 10 a Reasonable Model?

Test the following hypothesis

H_0 = "the dataset is generated by a state of rank at most r"
H_1 = "the dataset is generated by a state of rank higher than r"

Define the Pearson χ^2-statistic

$$T(r) = \sum_{\mathbf{s},\mathbf{d}} \frac{(N(\mathbf{s}|\mathbf{d}) - E(\mathbf{s}|\mathbf{d}))^2}{E(\mathbf{s}|\mathbf{d})}, \tag{2}$$

where $N(\mathbf{s}|\mathbf{d})$ are the counts from the real data.

Under the hypothesis H_0,

$$T(r) \sim \chi^2$$

distribution with number of degrees of freedom equal to the number of free parameters of the dataset minus the number of parameters of the model

$$\mathrm{df(r)} := 3^4 \cdot (2^4 - 1) - \mathrm{p(r, 4)}.$$

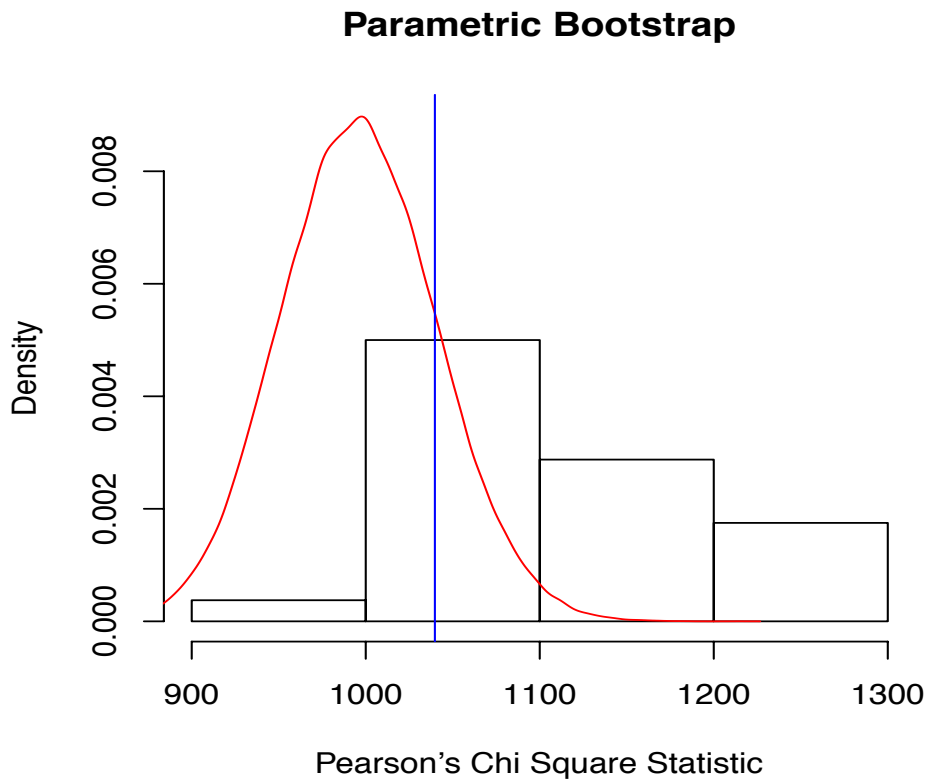
In this case, $r = 10$.

Is the State of Rank 10 a Reasonable Model?

In practice the χ^2 approximation works well for pure states ($r = 1$), and small rank states which have only a few small eigenvalues.

However, if the state has a significant number of small eigenvalues, the distribution of $T(r)$ may differ significantly from the asymptotic χ^2 distribution.

We employ *Bootstrap* techniques to estimate the distribution of $T(r)$ and then perform the test with respect to that distribution.



Conclusions and Outlook

Conclusions

- We proposed model selection as a general principle for approaching state estimation problems.
- The aim is to reduce the dimensionality of the problem by taking advantage of the "sparsity" properties of quantum states in realistic experimental settings.
- The philosophy is to try to find the simplest, or most *parsimonious* explanation of the data ...
- ... by fitting different models (often of increasing complexity) and choosing the estimator with the best trade-off between likelihood and complexity: **AIC and BIC**.
- Model selection gives sensible answers and can be used to used as an alternative to full tomography and compressed sensing.

Outlook

- The drawback is the computational complexity of finding the MLE over states of fixed rank.
- Combine ideas from different methods to come up with a fast, scalable and statistically efficient estimator.
- A possible future direction is apply model selection to state estimation for other types of models (e.g. MPS)
- Another topic of interest is the computation of honest and meaningful confidence intervals ("error-bars").
- There is need for a deeper theoretical understanding of the quantum tomography statistical model (e.g. effect of state's proximity to the boundary and asymptotics?)
- This would hopefully lead to improved estimation algorithms and model selection methods.

Acknowledgements

- Prof Rainer Blatt's Group (University of Innsbruck)
  - Thomas Monz and Phillip Schindler

  for many sharing with us their experimental data, fruitful discussions and their hospitality during our visits to Innsbruck.

- EPSRC: EP/E052290/1

- University of Nottingham Additional Sponsorship grant EP/J501499/1.

More details/results can be found here

`http://arxiv.org/abs/1206.4032`

Parameterising Models with Fixed Rank

We parameterize the density matrix ρ as

$$\rho = T^\star T$$

where T is an upper triangular matrix with strictly positive diagonal elements and parametrized by the vector of real numbers $\theta := (R, I, D) \in \mathbb{R}^{d^2-1}$ with

$$\begin{cases} R & := & (\mathrm{Re}(T_{12}), \ldots, \mathrm{Re}(T_{d-1,d})) \\ I & := & (\mathrm{Im}(T_{12}), \ldots, \mathrm{Im}(T_{d-1,d})) \\ D & := & (T_{22}, \ldots, T_{dd}) \end{cases} \tag{3}$$

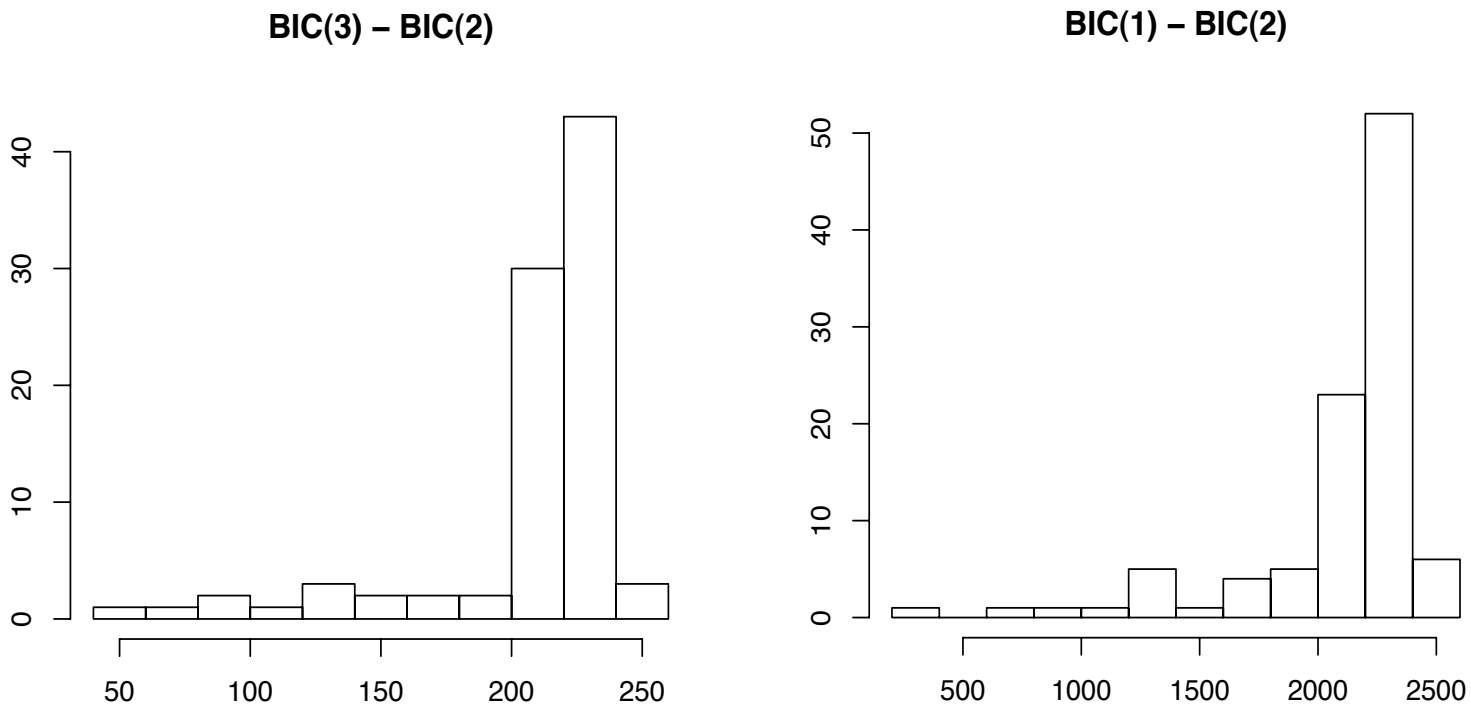
such that R, I are the real and imaginary parts of the off-diagonal elements ordered from the first to the $d-1$ row, and from left to right along each row.

θ must satisfy the constraints $D > 0$ and $\|R\|^2 + \|I\|^2 + \|D\|^2 < 1$, and the left-top element of T is equal to

$$T_{11} = T_{11}(\theta) = (1 - \|R\|^2 + \|I\|^2 + \|D\|^2)^{1/2} > 0.$$

Study 1: Randomly Chosen Low Rank States

While at first sight this may appear to be a surprisingly good performance, we can show that it agrees very well with the predictions of asymptotic theory.



Histograms of BIC differences when the true state is of rank 2

Study 2: One Ion simulations

		Measurement Repetitions				
		10	50	100	250	500
State 1	BIC	987	990	994	992	996
	AIC	945	944	919	927	930
State 2	BIC	25	83	183	394	706
	AIC	77	312	502	802	942
State 3	BIC	384	973	998	997	988
	AIC	594	992	998	997	988

- Pure state: both criteria require a small number of repetitions (of the order of 50) to give the correct answer.
- Almost pure state: for small n the difference between the log-likelihoods does not off-set the complexity penalty and both criteria choose rank one; at $n = 500$ the balance tips in favour of the rank 2 model, with AIC switching faster than BIC, on average.