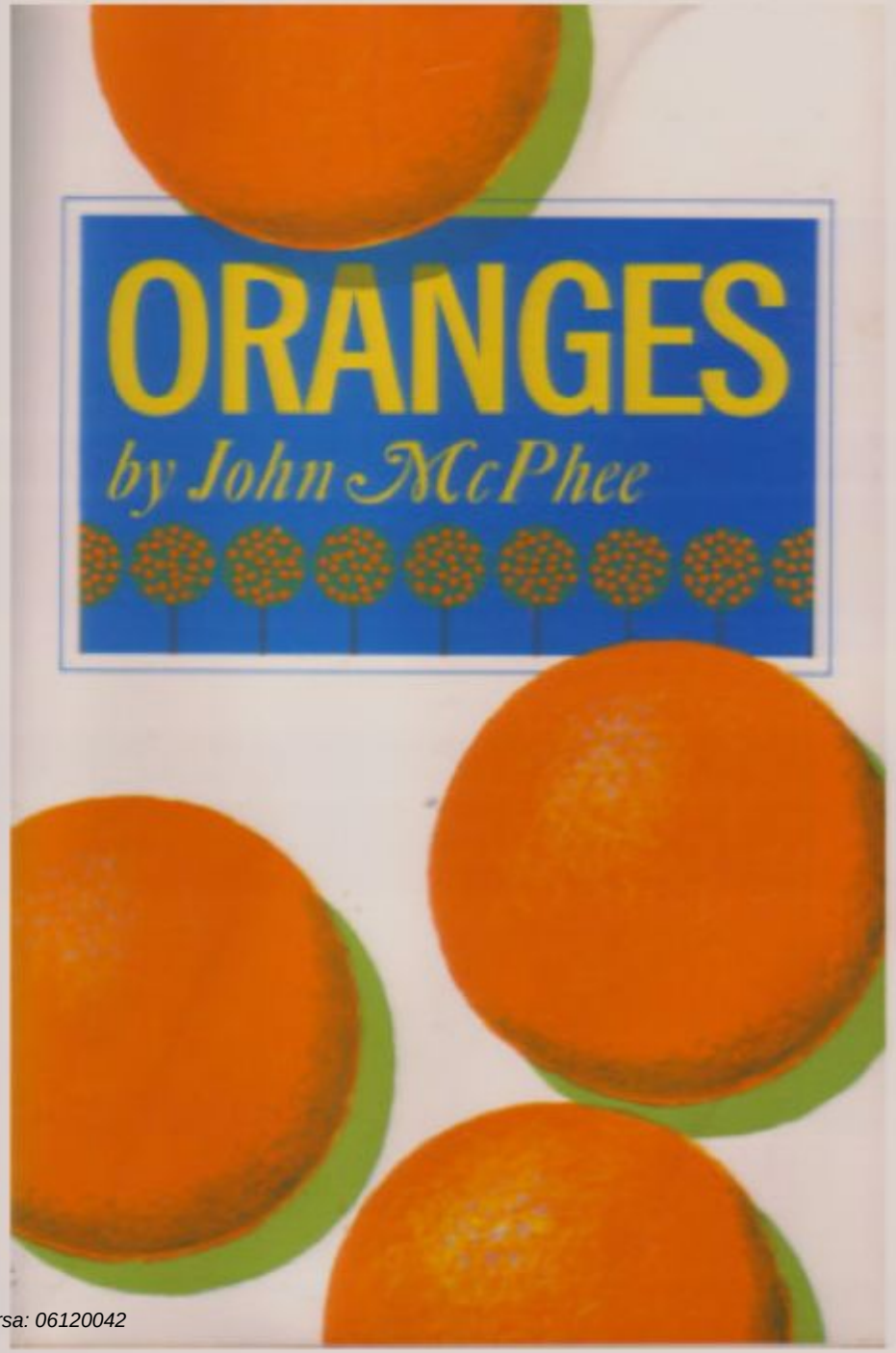
Title: Reliable Quantum State Estimation from Quantum Scoring Rules

Date: Dec 08, 2006 11:05 AM

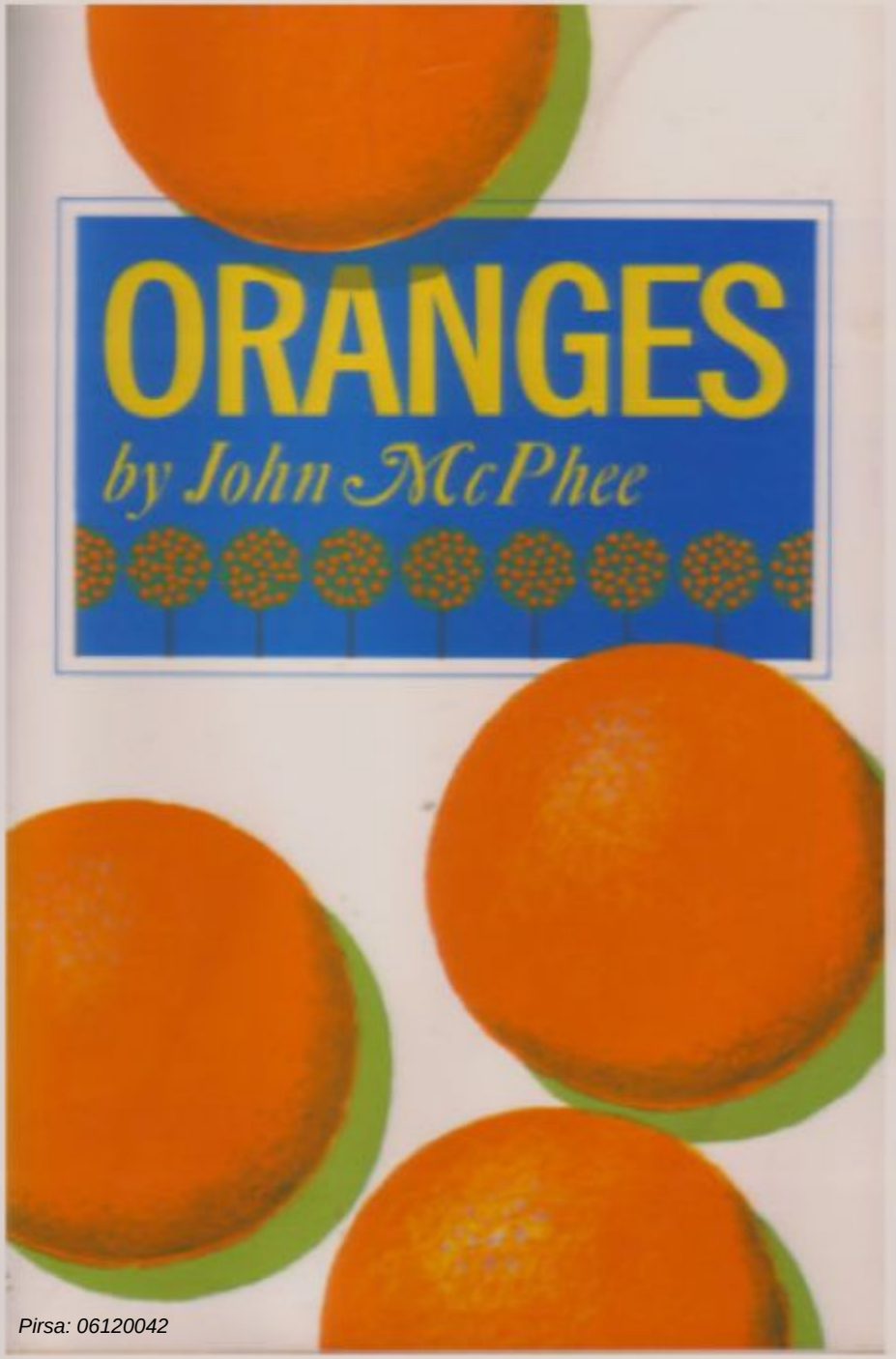
URL: http://pirsa.org/06120042

Abstract: Inferring a quantum system\'s state, from repeated measurements, is critical for verifying theories and designing quantum hardware. It\'s also surprisingly easy to do wrong, as illustrated by maximum likelihood estimation (MLE), the current state of the art. I\'ll explain why MLE yields unreliable and rank-deficient estimates, why you shouldn\'t be a quantum frequentist, and why we need a different approach. I\'ll show how operational divergences -- well-motivated metrics designed to evaluate estimates -- follow from quantum strictly proper scoring rules. This motivates Bayesian Mean Estimation (BME), and I\'ll show how it fixes most of the problems with MLE. I\'ll conclude with a couple of speculations about the future of quantum state and process estimatio



Oranges (Paperback)
by John McPhee
★★★★★ (18 customer reviews)

List Price: ~~$13.00~~
Price: **$10.40** & eligible for **FREE Super Saver Shipping** on orders over $25. Details
You Save: $2.60 (20%)



Oranges (Paperback)
by John McPhee
★★★★★★ (18 customer reviews)

List Price: ~~$13.00~~
Price: **$10.40** & eligible for **FREE Super Saver Shipping** on orders over $25. Details
You Save: $2.60 (20%)

Institute for Quantum Information

Reliable Quantum State Estimation

quant-ph/0611080

from

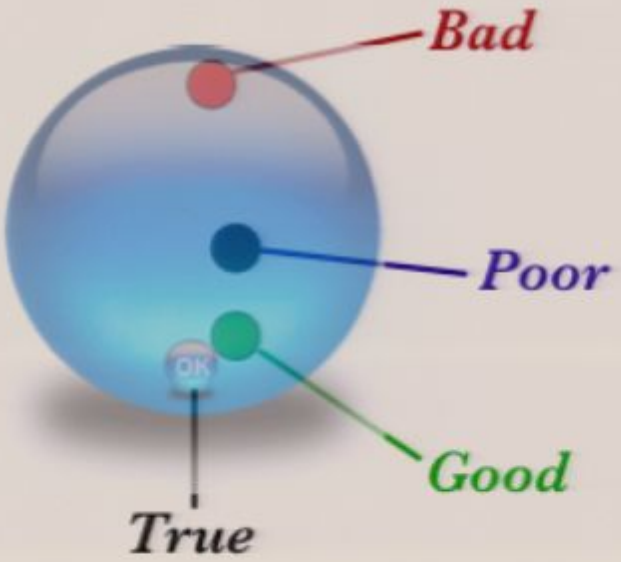
Quantum Scoring Rules

quant-ph/0603116

Robin Blume-Kohout (Caltech-IQI)

with Patrick Hayden (McGill)
and Karan Malhotra (IIT-Kanpur)

What is State Estimation?



- **Goal**: Characterize a *source* of quantum systems.
- You measure *N* identical copies.
- Measurement record:

$$\mathcal{M} = \left\{ \hat{E}_1, \hat{E}_2, \ldots \hat{E}_N \right\}$$

- Then report your best guess for ρ.
- **Key Application**: Verifying quantum hardware for fault tolerant quantum computation.

=> probabilities in $\left[10^{-3} \ldots 10^{-5}\right]$ are important!

A few points up front:

A few points up front:

- Estimation procedures consist of:
  a) a rule for making measurements,
  b) a rule for analyzing the data.

A few points up front:

- Estimation procedures consist of:
  a) a rule for making measurements,
  b) a rule for analyzing the data.

- I'm only talking about the *data analysis* part.

A few points up front:

- Estimation procedures consist of:
  a) a rule for making measurements,
  b) a rule for analyzing the data.

- I'm only talking about the *data analysis* part.

- Every [sane] procedure works roughly the same as $N \Rightarrow \infty$. I'm interested in $1 < N < \infty$.

What I'm a-gonna say

1. **Motivation**: problems with MLE*.
2. **Foundation**: quantum scoring rules.
3. **Solution**: the BME algorithm (& its features)
4. **Testing**: dueling procedures! (numerics)

* Maximum Likelihood Estimation -- the current state of the art.

What I'm a-gonna say

1. **Motivation**: problems with MLE*.
2. **Foundation**: quantum scoring rules.
3. **Solution**: the BME algorithm (& its features)
4. **Testing**: dueling procedures! (numerics)

* Maximum Likelihood Estimation -- the current state of the art.

What's the Problem?

- Current technology: **M**aximum **L**ikelihood **E**stimation

$$\mathcal{M} = \left\{\hat{E}_1, \hat{E}_2, \ldots \hat{E}_N\right\} \longrightarrow \mathcal{L}(\rho) \equiv p(\mathcal{M}|\rho) \longrightarrow \rho_{\mathrm{MLE}}$$

- ρ_{MLE} does not honestly represent the experimentalist's knowledge about the observed ensemble.
- **Why?** ρ_{MLE} typically has zero eigenvalues:

$$\lambda_i = \langle\phi_i| \rho_{\mathrm{MLE}} |\phi_i\rangle = 0$$

- Zero *eigenvalue* = zero *probability*

What's the Problem?

- Current technology: **M**aximum **L**ikelihood **E**stimation

$$\mathcal{M} = \left\{\hat{E}_1, \hat{E}_2, \ldots \hat{E}_N\right\} \longrightarrow \mathcal{L}(\rho) \equiv p(\mathcal{M}|\rho) \longrightarrow \rho_{\mathrm{MLE}}$$

- ρ_{MLE} does not honestly represent the experimentalist's knowledge about the observed ensemble.
- **Why?** ρ_{MLE} typically has zero eigenvalues:

$$\lambda_i = \langle\phi_i| \rho_{\mathrm{MLE}} |\phi_i\rangle = 0$$

- Zero *eigenvalue* = zero *probability*
  - ↳ absolute certainty
    - ↳ doesn't admit error bars!

Yes, it's a problem

PHYSICAL REVIEW A, VOLUME 64, 052312

Measurement of qubits

Daniel F. V. James,[1,*] Paul G. Kwiat,[2,3] William J. Munro,[4,5] and Andrew G. White[2,4]
[1]Theoretical Division T-4, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
[2]Physics Division P-23, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
[3]Department of Physics, University of Illinois, Urbana Champaign, Illinois 61801
[4]Department of Physics, University of Queensland, Brisbane, Queensland 4072, Australia
[5]Hewlett-Packard Laboratories, Filton Road, Stoke Gifford, Bristol BS34 8QZ, United Kingdom
(Received 20 March 2001; published 16 October 2001)

$$\hat{\rho} = \begin{pmatrix} 0.5069 & -0.0239+i0.0106 & -0.0412-i0.0221 & 0.4833+i0.0329 \\ -0.0239-i0.0106 & 0.0048 & 0.0023+i0.0019 & -0.0296-i0.0077 \\ -0.0412+i0.0221 & 0.0023-i0.0019 & 0.0045 & -0.0425+i0.0192 \\ 0.4833-i0.0329 & -0.0296+i0.0077 & -0.0425-i0.0192 & 0.4839 \end{pmatrix}$$

This matrix is illustrated in Fig. 3 (right). In this case, the matrix has eigenvalues 0.986 022, 0.013 977 7, 0, and 0; and $\mathrm{Tr}\{\hat{\rho}^2\} = 0.972\,435$, indicating that, while the linear reconstruction gave a nonphysical density matrix, the maximum likelihood reconstruction gives a legitimate density matrix.

1. D. F. V. James et. al., "Measurement of Qubits," *Phys. Rev. A*, 64:052312 (**2001**)

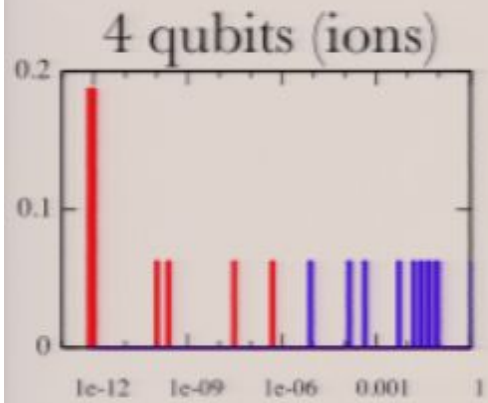
Yes, it's a problem

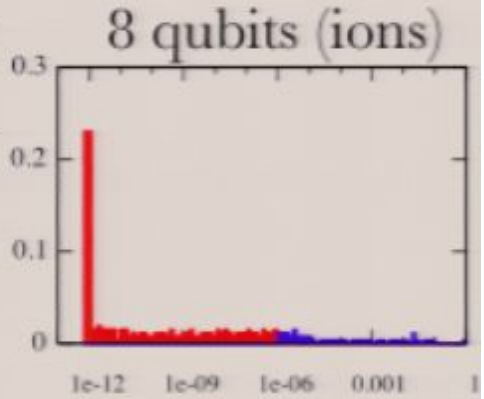
Vol 438|1 December 2005|doi:10.1038/nature04279 nature

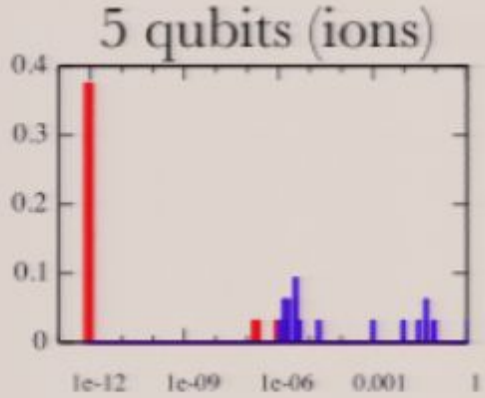
LETTERS

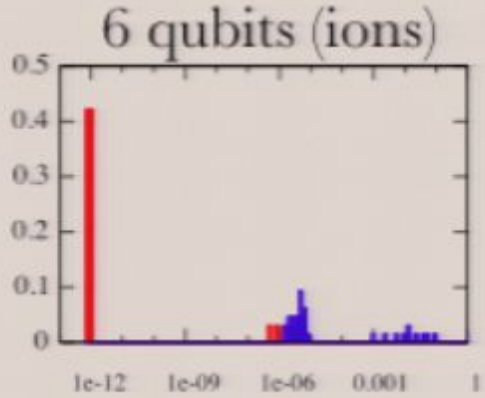
Scalable multiparticle entanglement of trapped ions

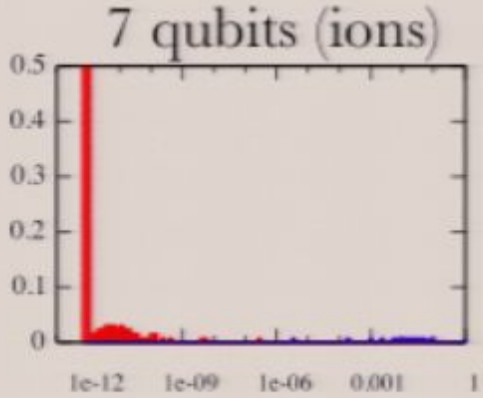
H. Häffner[1,2], W. Hänsel[1], C. F. Roos[1,2], J. Benhelm[1,2], D. Chek-al-kar[1], M. Chwalla[1], T. Körber[1,2], U. D. Rapol[1,2], M. Riebe[1], P. O. Schmidt[1], C. Becher[1]†, O. Gühne[3], W. Dür[2,3] & R. Blatt[1,2]







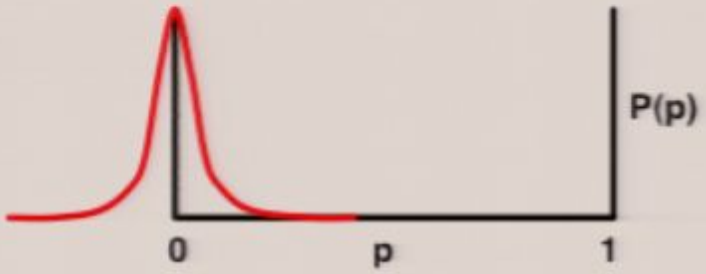




1. D. F. V. James et. al., "Measurement of Qubits," *Phys. Rev. A*, 64:052312 (**2001**)
2. Häffner et. al., "Scalable multiparticle entanglement of trapped ions," *Nature*, 438:643-6 (**2005**)

Error Bars?

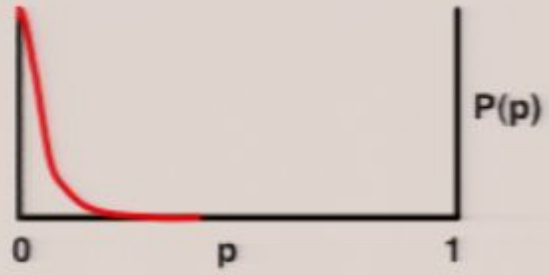
- "But aren't there error bars in there?"
- Not always.
- What does $p = 0 \pm 0.1$ mean?



Error Bars?

- "But aren't there error bars in there?"
- Not always.
- What does ~~$p = 0 \pm 0.1$~~ mean?

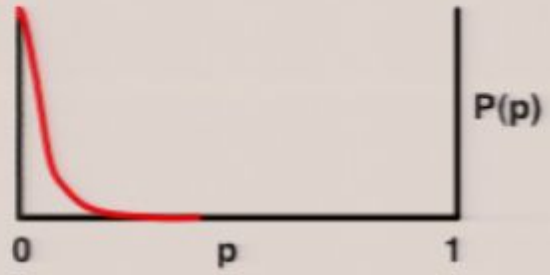
$$\Longrightarrow p = 0.05 \pm$$



Error Bars?

- "But aren't there error bars in there?"
- Not always.
- What does ~~$p = 0 \pm 0.1$~~ mean?

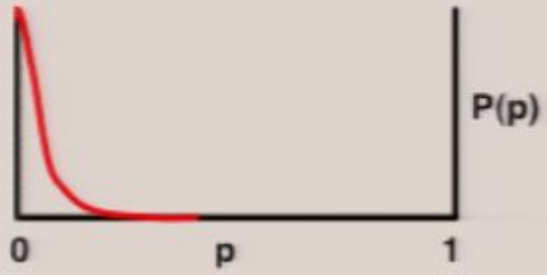
$$\Longrightarrow p = 0.05 \pm 0.05$$



Error Bars?

- "But aren't there error bars in there?"
- Not always.
- What does ~~$p = 0 \pm 0.1$~~ mean?

  $\Longrightarrow p = 0.05 \pm 0.05$

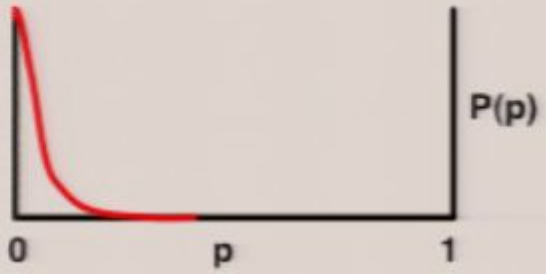


- Increasing *small* probabilities
  => decreasing *large* ones

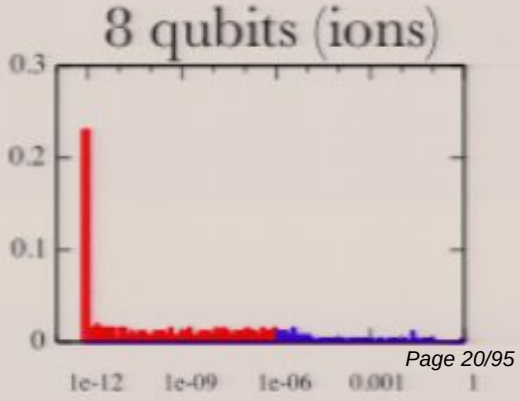
Error Bars?

- "But aren't there error bars in there?"
- Not always.
- What does ~~$p = 0 \pm 0.1$~~ mean?

  $\Longrightarrow p = 0.05 \pm 0.05$



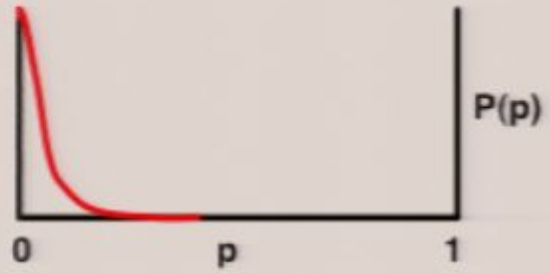
- Increasing *small* probabilities
  => decreasing *large* ones
- Effect of "fixing"
  ~218 zero eigenvalues?

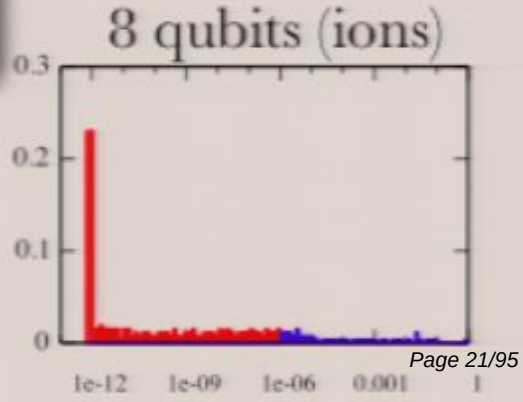


Error Bars?

$$\rho = \begin{pmatrix} .75 & & & & \\ & .06 & & & \\ & & \ddots & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}$$

ere?"





- Effect of "fixing"
  ~218 zero eigenvalues?

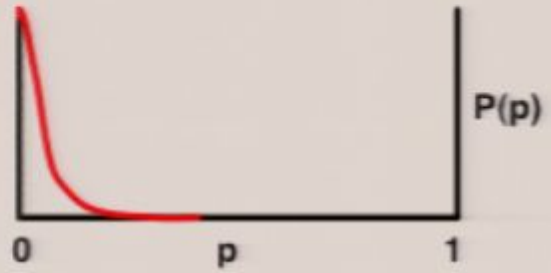
Error Bars?

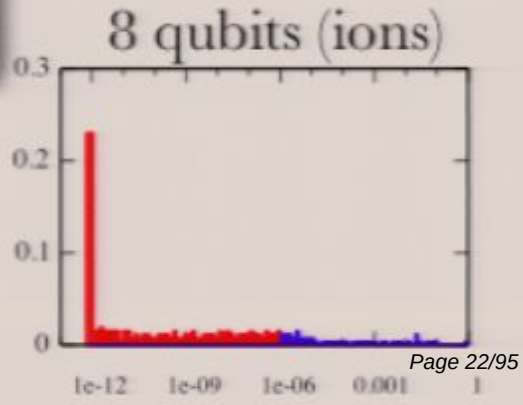
$$\rho = \begin{pmatrix} .75 & & & & \\ & .06 & & & \\ & & \ddots & & \\ & & & \varepsilon & \\ & & & & \ddots \\ & & & & & \varepsilon \end{pmatrix}$$

7×10^5	measurements
$65,536$	indep. components
$6,561$	different meas.
256	eigenvalues

$\varepsilon = \varepsilon(N_{\text{measurements}})$ . . . what is N?

ere?"





- Effect of "fixing"
  ~218 zero eigenvalues?

Error Bars?

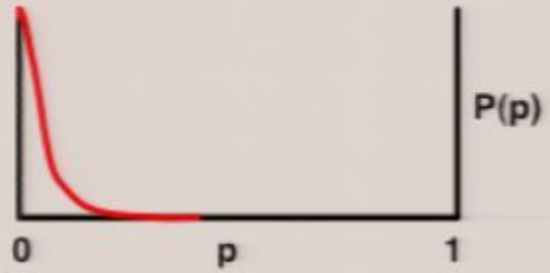
$$\rho = \begin{pmatrix} .75 & & & & \\ & .06 & & & \\ & & \ddots & & \\ & & & 10^{-6} & \\ & & & & \ddots \\ & & & & & 10^{-6} \end{pmatrix}$$

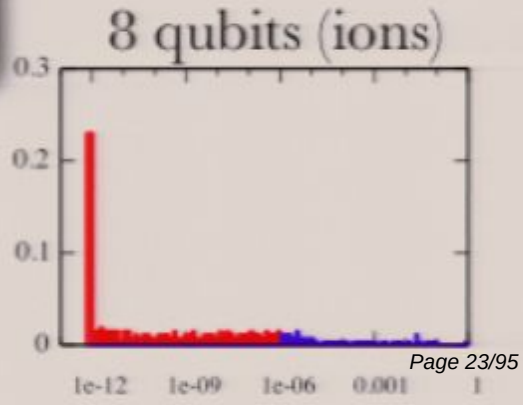
7×10^5	measurements
$65,536$	indep. components
$6,561$	different meas.
256	eigenvalues

$\varepsilon = \varepsilon(N_{\text{measurements}})$ . . . what is N?

$N = 7 \times 10^5$?

ere?"



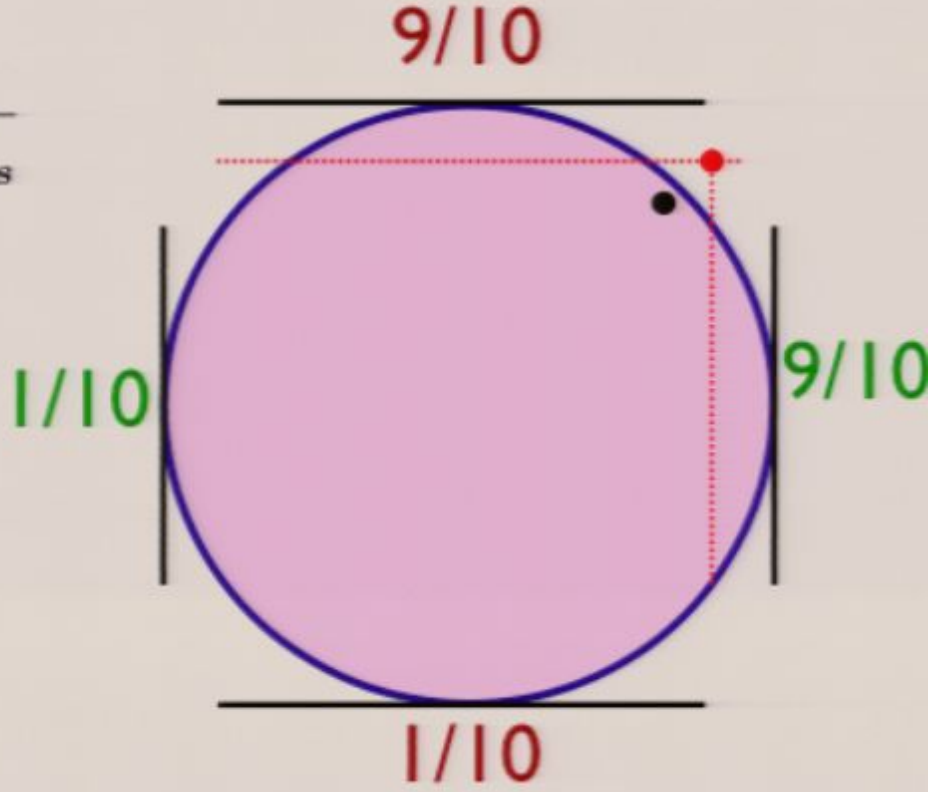


- Effect of "fixing"
  ~218 zero eigenvalues?


So... how did
we get here?

Tomography

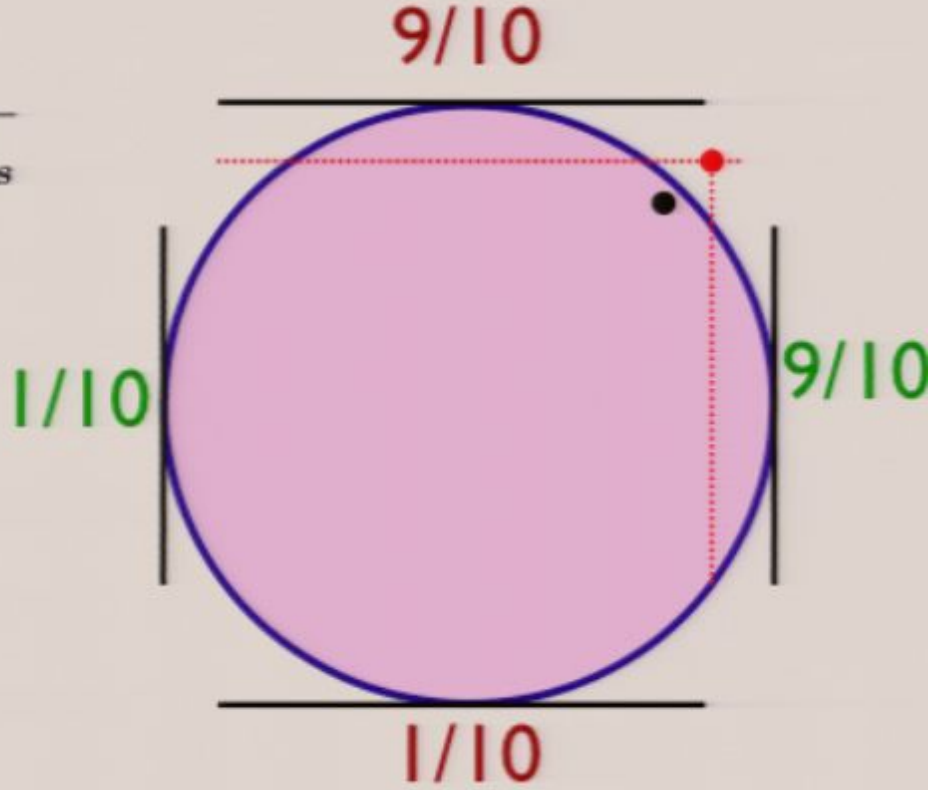
- Assume $p\left(\hat{M}_i\right) = \frac{n_i}{N_{meas}}$
- Probabilities give expectation values.
- Expectation values uniquely identify the density matrix.
- Negativity: $\rho_{\text{tomo}} \ngeq 0$



Why? (a) equation of frequencies with probabilities,
(b) observables aren't truly independent.

Tomography

- Assume $p\left(\hat{M}_i\right) = \dfrac{n_i}{N_{meas}}$
- Probabilities give expectation values.
- Expectation values uniquely identify the density matrix.
- Negativity: $\rho_{\text{tomo}} \ngeq 0$



Why? (a) equation of frequencies with probabilities,
(b) observables aren't truly independent.

Maximum Likelihood (MLE)

- Report the state that *makes the data most likely* -- i.e., maximize the likelihood of data.

$$\mathcal{L}(\rho) = p(\mathcal{M}|\rho) = \mathrm{Tr}\left(\hat{M}_1\rho\right)\mathrm{Tr}\left(\hat{M}_2\rho\right)\ldots\mathrm{Tr}\left(\hat{M}_N\rho\right)$$

- Involves maximization over valid states (parameterized by their square roots)

- Z. Hradil, "Quantum-state Estimation," *Phys. Rev. A*, 55:R1561 (**1997**)

How MLE Works

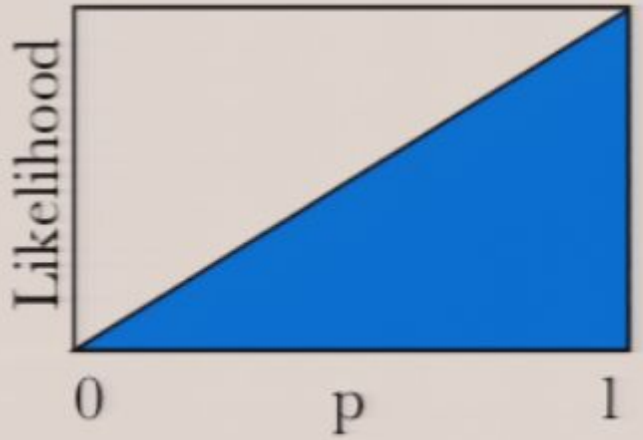
How MLE Works

Example 1: Bias of a Coin

How MLE Works

Example 1: Bias of a Coin

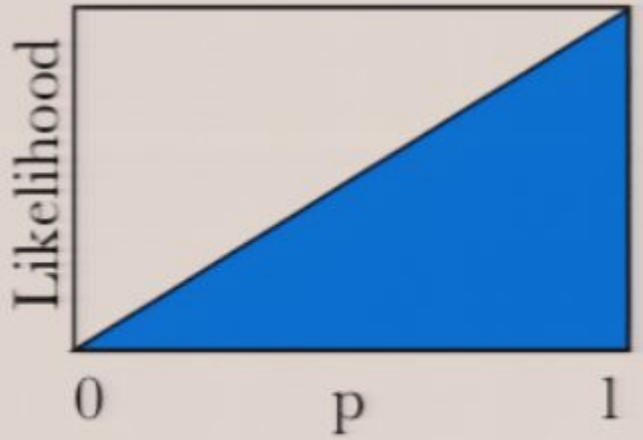
(a)
1 × head
0 × tails



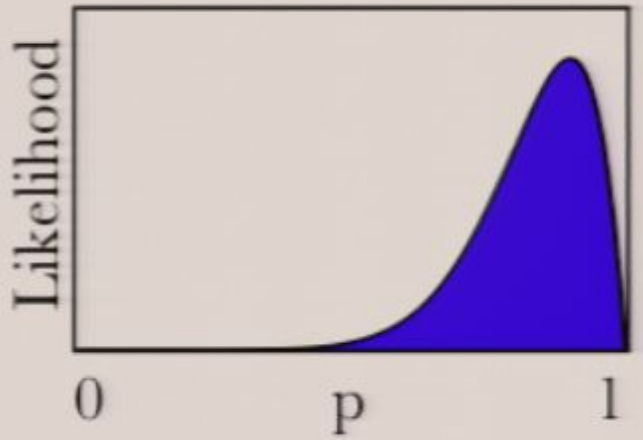
How MLE Works

Example 1: Bias of a Coin

(a)
1 × head
0 × tails



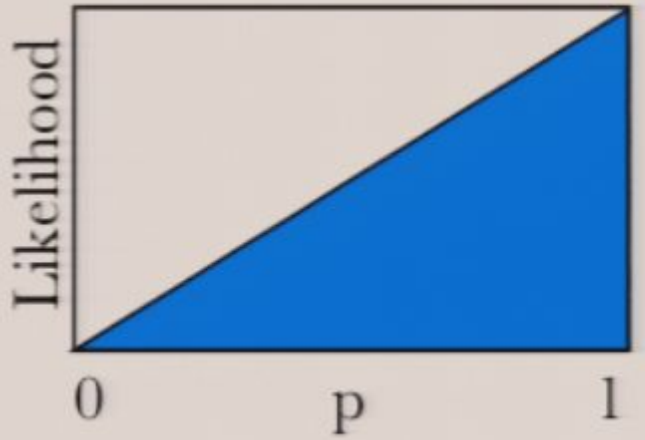
(b)
9 × heads
1 × tails



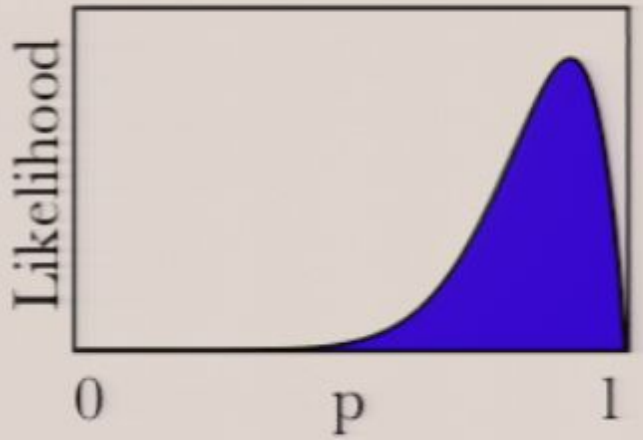
How MLE Works

Example 1: Bias of a Coin

(a)
1 × head
0 × tails



(b)
9 × heads
1 × tails

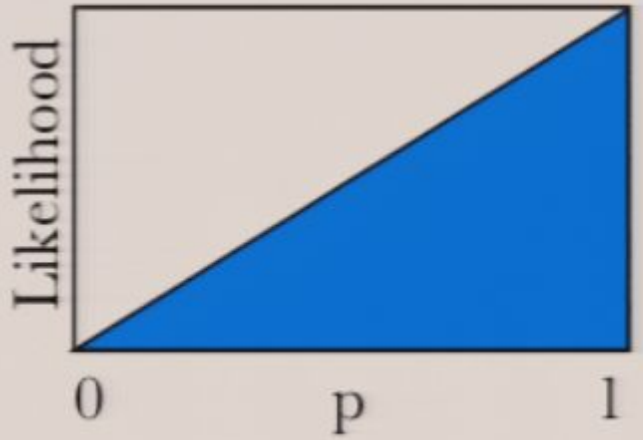


Example 2: Qubit

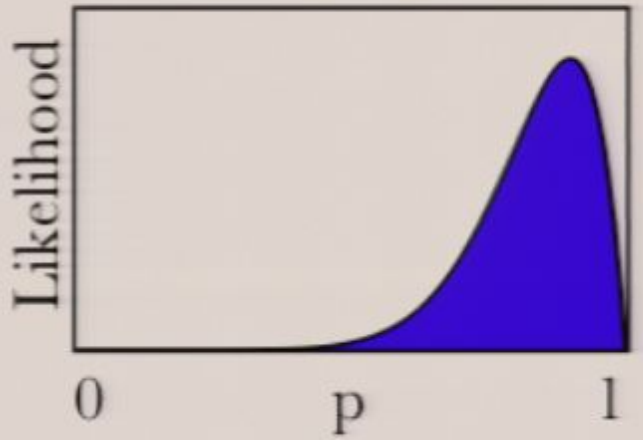
How MLE Works

Example 1: Bias of a Coin

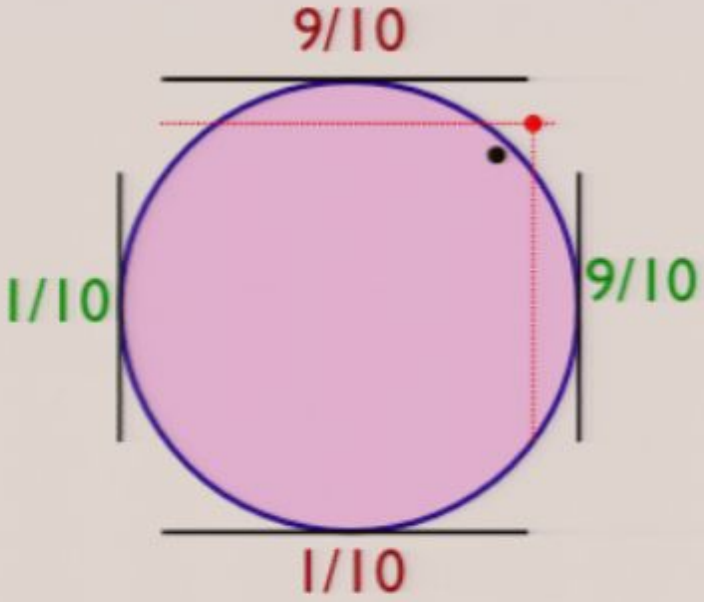
(a)
1 × head
0 × tails



(b)
9 × heads
1 × tails



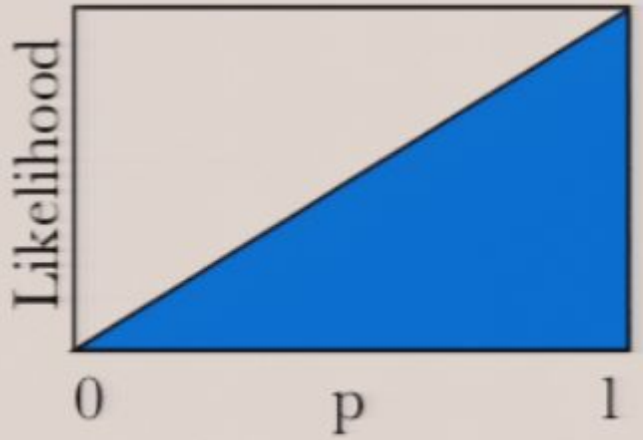
Example 2: Qubit



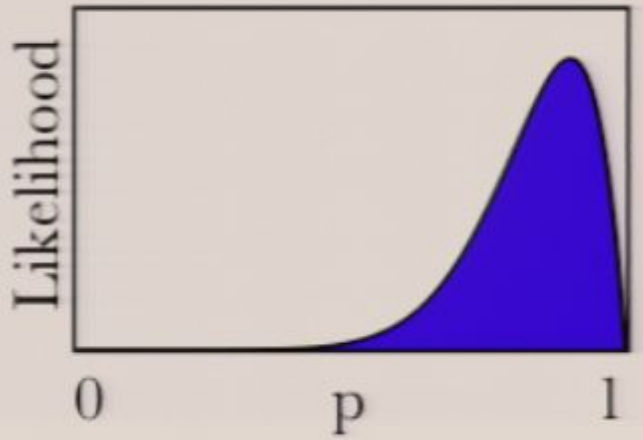
How MLE Works

Example 1: Bias of a Coin

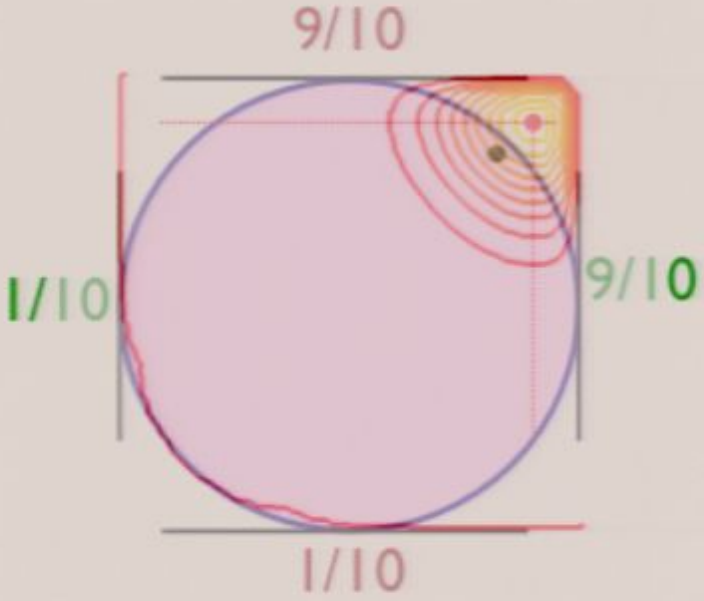
(a)
1 × head
0 × tails



(b)
9 × heads
1 × tails



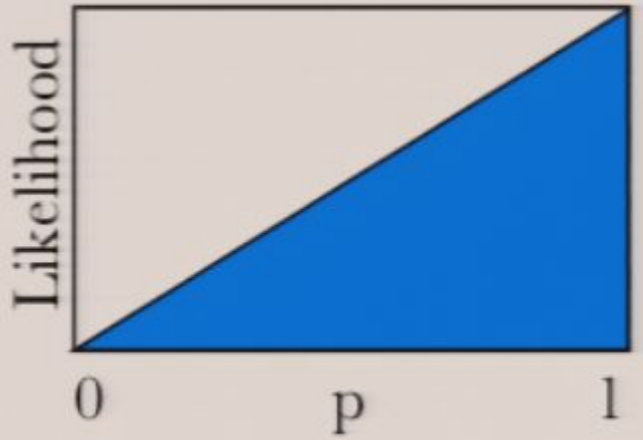
Example 2: Qubit



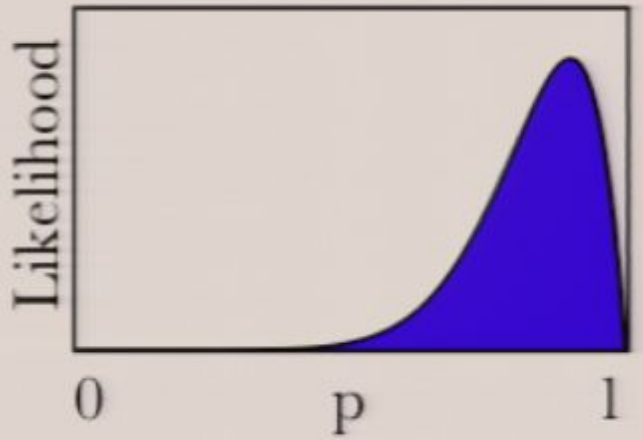
How MLE Works

Example 1: Bias of a Coin

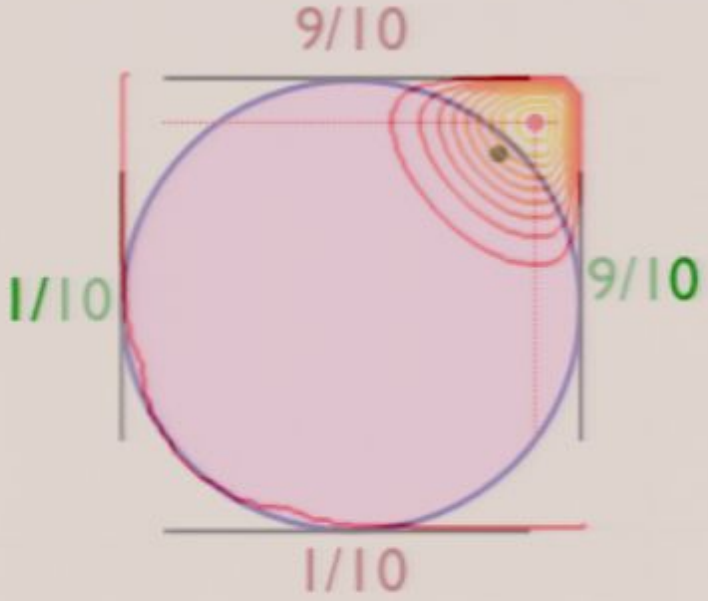
(a)
1 × head
0 × tails



(b)
9 × heads
1 × tails



Example 2: Qubit



MLE replaces *negative* eigenvalues with *zeros*.

A minimal fix for tomography.

What's going wrong?

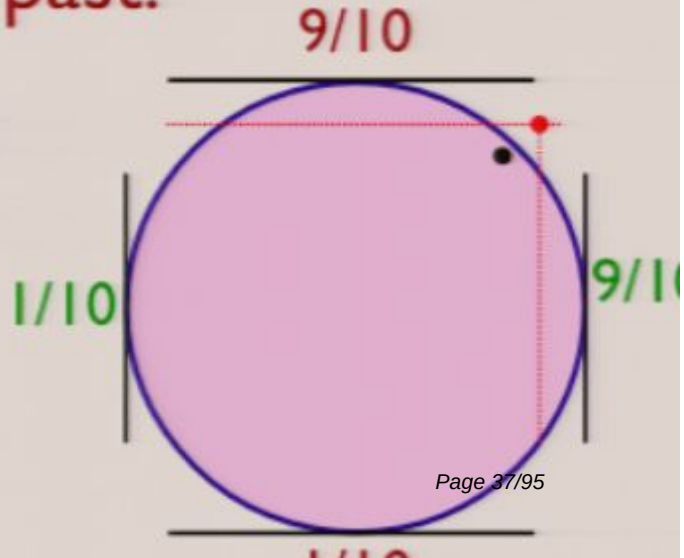
(the really, *really* short version!)

- MLE is a great way of *fitting the data*.
- Best estimate: *predicts* what we've seen.
- MLE takes observed *frequencies* literally -- "The future will look like the past."
- But... in QM the future might contain measurements that we haven't made yet!

What's going wrong?

(the really, *really* short version!)

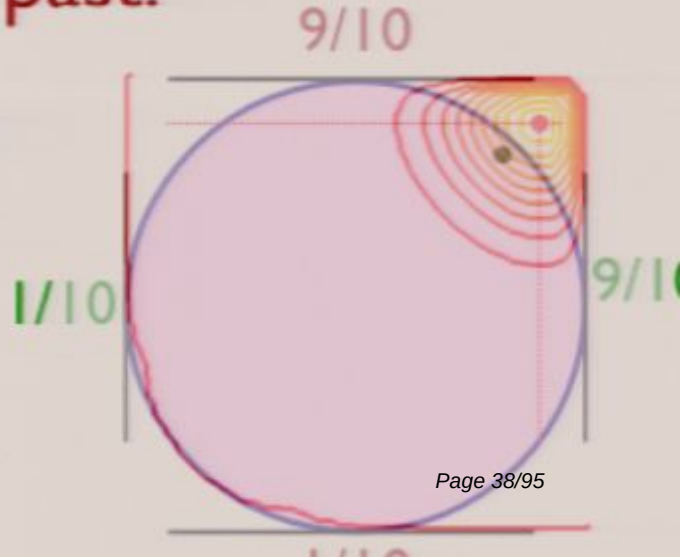
- MLE is a great way of *fitting the data*.
- Best estimate: *predicts* what we've seen.
- MLE takes observed *frequencies* literally -- "The future will look like the past."
- But... in QM the future might contain measurements that we haven't made yet!



What's going wrong?

(the really, *really* short version!)

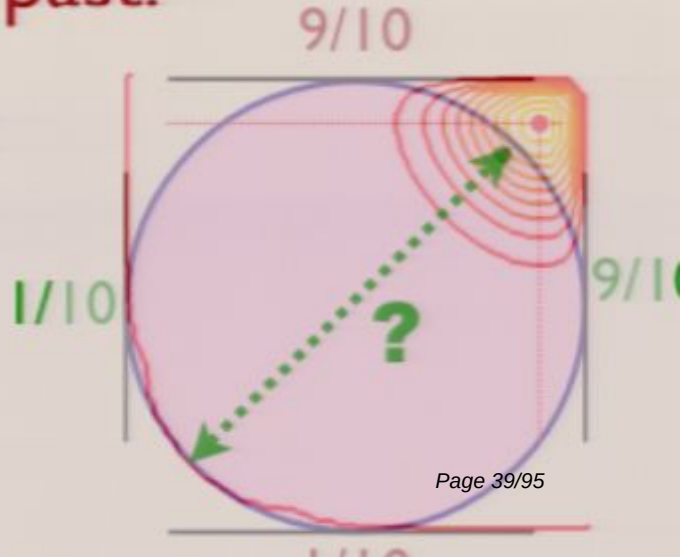
- MLE is a great way of *fitting the data*.
- Best estimate: *predicts* what we've seen.
- MLE takes observed *frequencies* literally -- "The future will look like the past."
- But... in QM the future might contain measurements that we haven't made yet!



What's going wrong?

(the really, *really* short version!)

- MLE is a great way of *fitting the data*.
- Best estimate: *predicts* what we've seen.
- MLE takes observed *frequencies* literally -- "The future will look like the past."
- But... in QM the future might contain measurements that we haven't made yet!



...so MLE assigns p=0 to events that were never observed not to happen...

Hmm... what to do?

1. **Motivation**: problems with MLE.
2. **Foundation**: quantum scoring rules.
3. **Solution**: the BME algorithm (& its features)
4. **Testing**: dueling procedures! (numerics)

Hmm... what to do?

1. **Motivation**: problems with MLE.

2. **Foundation**: quantum scoring rules.

3. **Solution**: the BME algorithm (& its features)

4. **Testing**: dueling procedures! (numerics)

States as Predictions

- Quantum states are like probability distributions: they predict the outcome of future measurements.
- Lets define a metric $f(\rho : \sigma)$ that measures how well σ predicts measurements on ρ.

  1. **The best estimate of ρ is ρ itself.**
     – i.e., $f(\rho : \rho) > f(\rho : \sigma)$ for all $\sigma \neq \rho$.

  2. **$f(\rho : \sigma)$ should correspond to an *operational* test**
     – i.e., someone's utility (reward or cost) for some practical procedure.

States as Predictions

- Quantum states are like probability distributions: they predict the outcome of future measurements.
- Lets define a metric $f(\rho : \sigma)$ that measures how well σ predicts measurements on ρ.

  1. **The best estimate of ρ is ρ itself.**
     – i.e., $f(\rho : \rho) > f(\rho : \sigma)$ for all $\sigma \neq \rho$.

  2. **$f(\rho : \sigma)$ should correspond to an *operational* test**
     – i.e., someone's utility (reward or cost) for some practical procedure.

$\Longrightarrow$ **Quantum Strictly Proper Scoring Rules**

Quantum
Strictly Proper Scoring Rules

Quantum
Strictly Proper Scoring Rules

Victor the Verifier measures a copy of ρ, and pays you $R_i(\sigma)$ if he gets outcome i.

Quantum
Strictly Proper Scoring Rules

Victor the Verifier measures a copy of ρ, and pays you $R_i(\sigma)$ if he gets outcome i.

Victor's goal: to motivate you to tell the truth *if you know ρ exactly.*

Your goal: to maximize your expected reward (you greedy Scrooge, you!)

Quantum
Strictly Proper Scoring Rules

Victor the Verifier measures a copy of ρ, and pays you $R_i(\sigma)$ if he gets outcome i.

Define $f(\rho : \sigma)$ as your expected reward, from Victor, for reporting σ when the true state is ρ.

Quantum
Strictly Proper Scoring Rules

Victor the Verifier measures a copy of ρ, and pays you $R_i(\sigma)$ if he gets outcome i.

Define $f(\rho : \sigma)$ as your expected reward, from Victor, for reporting σ when the true state is ρ.

$$f(\rho : \sigma) = \sum_i p(i) R_i(\sigma)$$

Measurement $= \{E_i\}$: $\sum E_i = \mathbb{1}$.

$$f(\rho : \sigma) = \sum_i \mathrm{Tr}[\rho E_i] R_i(\sigma)$$
$$= \mathrm{Tr}[\rho \mathcal{R}(\sigma)]$$

where $\mathcal{R}(\sigma) = \sum_i E_i R_i(\sigma)$.

Define "value": $G(\rho) \equiv f(\rho : \rho) = \mathrm{Tr}[\rho \mathcal{R}(\rho)]$.

Quantum
Strictly Proper Scoring Rules

Victor the Verifier measures a copy of ρ, and pays you $R_i(\sigma)$ if he gets outcome i.

1) $f(\rho : \sigma) = \sum_i p(i) R_i(\sigma)$

Measurement $= \{E_i\} : \quad \sum E_i = \mathbb{1}$.

$$f(\rho : \sigma) = \sum_i \mathrm{Tr}[\rho E_i] R_i(\sigma)$$
$$= \mathrm{Tr}[\rho \mathcal{R}(\sigma)]$$

where $\mathcal{R}(\sigma) = \sum_i E_i R_i(\sigma)$.

Define "value": $G(\rho) \equiv f(\rho : \rho) = \mathrm{Tr}[\rho \mathcal{R}(\rho)]$.

Quantum
Strictly Proper Scoring Rules

Victor the Verifier measures a copy of ρ, and pays you $R_i(\sigma)$ if he gets outcome i.

Now, demand that the reward for truth be greater than the reward for *any* lie.

Quantum
Strictly Proper Scoring Rules

Victor the Verifier measures a copy of ρ, and pays you $R_i(\sigma)$ if he gets outcome i.

Now, demand that the reward for truth be greater than the reward for *any* lie.

2) $f(\rho : \rho) > f(\rho : \sigma)$ if $\sigma \neq \rho$.

$$f(\rho : \rho) > f(\sigma : \sigma) + f(\rho : \sigma) - f(\sigma : \sigma)$$

$$G(\rho) > G(\sigma) + \mathrm{Tr}\left[(\rho - \sigma)\mathcal{R}(\sigma)\right]$$

Quantum
Strictly Proper Scoring Rules

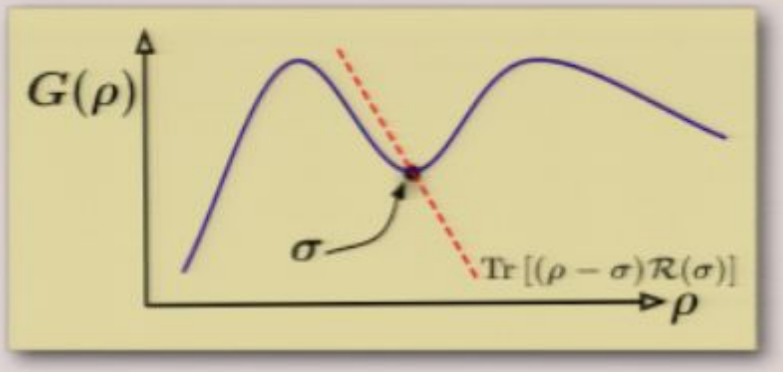
Victor the Verifier measures a copy of ρ, and pays you $R_i(\sigma)$ if he gets outcome i.

Now, demand that the reward for truth be greater than the reward for *any* lie.

2) $f(\rho : \rho) > f(\rho : \sigma)$ if $\sigma \neq \rho$.

$$f(\rho : \rho) > f(\sigma : \sigma) + f(\rho : \sigma) - f(\sigma : \sigma)$$

$$G(\rho) > G(\sigma) + \mathrm{Tr}\left[(\rho - \sigma)\mathcal{R}(\sigma)\right]$$



Quantum
Strictly Proper Scoring Rules

Victor the Verifier measures a copy of ρ, and pays you $R_i(\sigma)$ if he gets outcome i.

1) $f($

Meas

Now, demand that the reward for truth be greater than the reward for *any* lie.

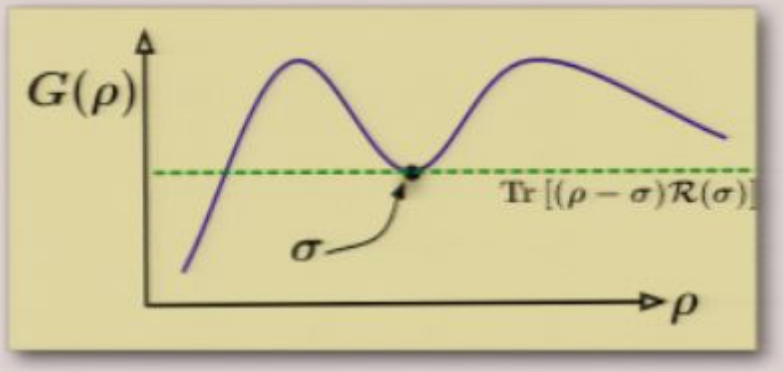
2) $f(\rho:\rho) > f(\rho:\sigma)$ if $\sigma \neq \rho$.

$$f(\rho:\rho) > f(\sigma:\sigma) + f(\rho:\sigma) - f(\sigma:\sigma)$$

$$G(\rho) > G(\sigma) + \mathrm{Tr}\left[(\rho-\sigma)\mathcal{R}(\sigma)\right]$$

where

Defin



- $f(\rho:\sigma)$ is a subtangent to $G(\rho)$

Quantum
Strictly Proper Scoring Rules

Victor the Verifier measures a copy of ρ, and pays you $R_i(\sigma)$ if he gets outcome i.

1) $f($

Meas

Now, demand that the reward for truth be greater than the reward for *any* lie.

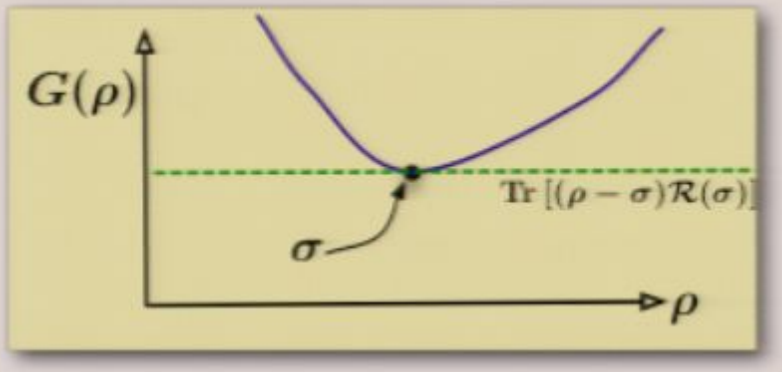
2) $f(\rho:\rho) > f(\rho:\sigma)$ if $\sigma \neq \rho$.

$$f(\rho:\rho) > f(\sigma:\sigma) + f(\rho:\sigma) - f(\sigma:\sigma)$$

$$G(\rho) > G(\sigma) + \mathrm{Tr}\left[(\rho-\sigma)\mathcal{R}(\sigma)\right]$$

where

Defin



- $f(\rho:\sigma)$ is a subtangent to $G(\rho)$
- $G(\rho)$ must be strictly convex.

Quantum
Strictly Proper Scoring Rules

Victor the Verifier measures a copy of ρ, and pays you $R_i(\sigma)$ if he gets outcome i.

1) $f(\rho:\sigma) = \sum_i p(i) R_i(\sigma)$

Measurement $= \{E_i\}: \quad \sum E_i = \mathbb{1}$.

$$f(\rho:\sigma) = \sum_i \mathrm{Tr}[\rho E_i] R_i(\sigma)$$
$$= \mathrm{Tr}[\rho \mathcal{R}(\sigma)]$$

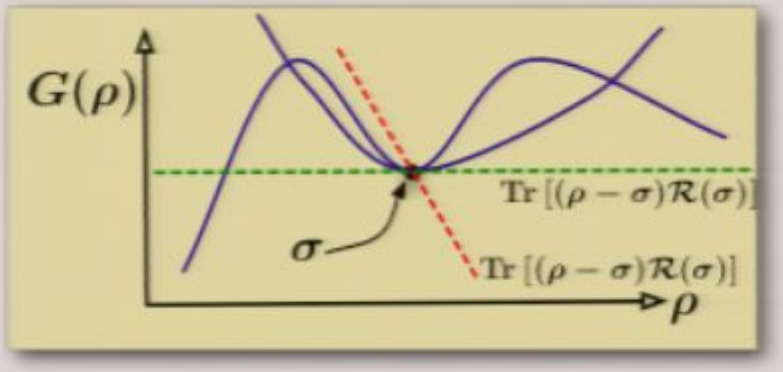
where $\mathcal{R}(\sigma) = \sum_i E_i R_i(\sigma)$.

Define "value": $G(\rho) \equiv f(\rho:\rho) = \mathrm{Tr}[\rho \mathcal{R}(\rho)]$.

2) $f(\rho:\rho) > f(\rho:\sigma)$ if $\sigma \neq \rho$.

$f(\rho:\rho) > f(\sigma:\sigma) + f(\rho:\sigma) - f(\sigma:\sigma)$

$G(\rho) > G(\sigma) + \mathrm{Tr}\left[(\rho-\sigma)\mathcal{R}(\sigma)\right]$



- $f(\rho:\sigma)$ is a subtangent to $G(\rho)$
- $G(\rho)$ must be strictly convex.

Quantum
Strictly Proper Scoring Rules

Victor the Verifier measures a copy of ρ, and pays you $R_i(\sigma)$ if he gets outcome i.

1) $f(\rho:\sigma) = \sum_i p(i) R_i(\sigma)$

Measurement $= \{E_i\}: \quad \sum E_i = \mathbb{1}$.

$$f(\rho:\sigma) = \sum_i \mathrm{Tr}[\rho E_i] R_i(\sigma)$$
$$= \mathrm{Tr}[\rho \mathcal{R}(\sigma)]$$

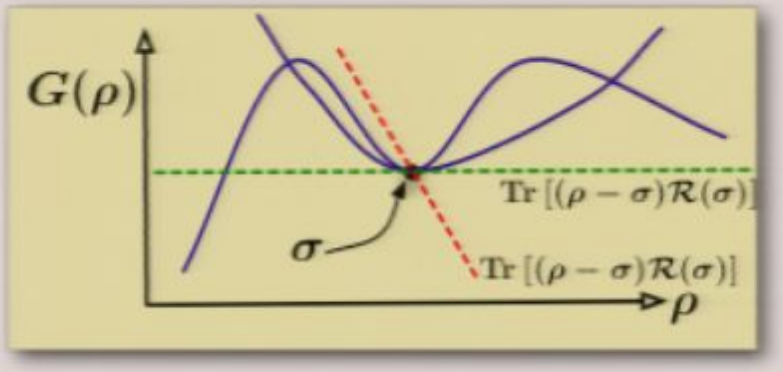
where $\mathcal{R}(\sigma) = \sum_i E_i R_i(\sigma)$.

Define "value": $G(\rho) \equiv f(\rho:\rho) = \mathrm{Tr}[\rho \mathcal{R}(\rho)]$.

2) $f(\rho:\rho) > f(\rho:\sigma)$ if $\sigma \neq \rho$.

$f(\rho:\rho) > f(\sigma:\sigma) + f(\rho:\sigma) - f(\sigma:\sigma)$

$G(\rho) > G(\sigma) + \mathrm{Tr}\left[(\rho-\sigma)\mathcal{R}(\sigma)\right]$



- $f(\rho:\sigma)$ is a subtangent to $G(\rho)$
- $G(\rho)$ must be strictly convex.

- The estimator's *expected loss* for lying is $\Delta(\rho:\sigma) \equiv G(\rho) - f(\rho:\sigma)$.
- $\Delta(\rho:\sigma)$ is an *operational divergence* = good measure of σ's predictive accuracy.
- Operational divergences are $1:1$ with strictly convex "entropies" $G(\rho)$.

A survey of metrics

- GOOD METRICS (OPERATIONAL DIVERGENCES)
  1. L_2 distance: $\mathrm{Tr}\left[(\rho-\sigma)^2\right] \Leftrightarrow R_i = 2\langle i|\,\sigma\,|i\rangle - \mathrm{Tr}(\sigma^2) - 1$
  2. Relative entropy: $S(\rho||\sigma) = \mathrm{Tr}\left[\rho \log \frac{\sigma}{\rho}\right] \Leftrightarrow R_i = \ln(\langle i|\,\sigma\,|i\rangle)$
- BAD METRICS
  1. Overlap: $1 - \mathrm{Tr}[\rho\sigma]$ — *Improper (motivates lying)*
  2. Trace-norm: $||\rho-\sigma||_1$
  3. Fidelity: $1 - \left[\mathrm{Tr}\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right]^2$

  } *Non-operational*

A survey of metrics

- GOOD METRICS (OPERATIONAL DIVERGENCES)
  1. L_2 distance: $\mathbf{Tr}\left[(\rho-\sigma)^2\right] \Leftrightarrow R_i = 2\langle i|\,\sigma\,|i\rangle - \mathbf{Tr}(\sigma^2) - 1$
  2. Relative entropy: $S(\rho||\sigma) = \mathbf{Tr}\left[\rho \log \frac{\sigma}{\rho}\right] \Leftrightarrow R_i = \ln(\langle i|\,\sigma\,|i\rangle)$
- BAD METRICS
  1. Overlap: $1 - \mathbf{Tr}[\rho\sigma]$ — *Improper (motivates lying)*
  2. Trace-norm: $||\rho-\sigma||_1$
  3. Fidelity: $1 - \left[\mathbf{Tr}\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right]^2$

} *Non-operational*

Operational divergences compare *physical states* to *classical descriptions*.

So... what do we do?

1. **Motivation**: problems with MLE.
2. **Foundation**: quantum scoring rules.
3. **Solution**: the BME algorithm (& its features)
4. **Testing**: dueling procedures! (numerics)

So... what do we do?

1. **Motivation**: problems with MLE.

2. **Foundation**: quantum scoring rules.

3. **Solution**: the BME algorithm (& its features)

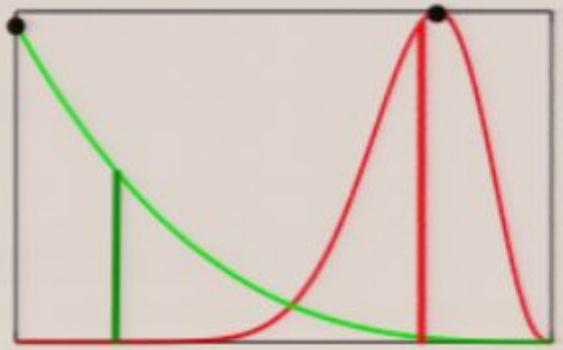
4. **Testing**: dueling procedures! (numerics)

Bayesian Mean Estimation

- Procedure: report the mean of a probability distribution obtained via Bayes' Rule.
- Begin with a *prior distribution*, $\pi_0(\rho)\mathrm{d}\rho$.
- Update to $\pi(\rho) = \mathcal{L}(\rho)\pi_0(\rho)$.
- Compute the mean value: $\hat{\sigma} = \int \hat{\rho}\, \pi(\rho)\mathrm{d}\rho$

MLE reports the ***mode***

BME reports the ***mean***



Bayesian Inference

1. Begin with a prior: $\boldsymbol{\pi_0(\rho)\mathrm{d}\rho}$.
2. Make $\boldsymbol{N}$ measurements
   $\Rightarrow$ record: $\mathbf{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\boldsymbol{\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho}$

$= \mathbf{Tr}(\hat{M}_1\rho)\mathbf{Tr}(\hat{M}_2\rho)\ldots\mathbf{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$

Example: spin-½

Measure Y, X (ignore Z)

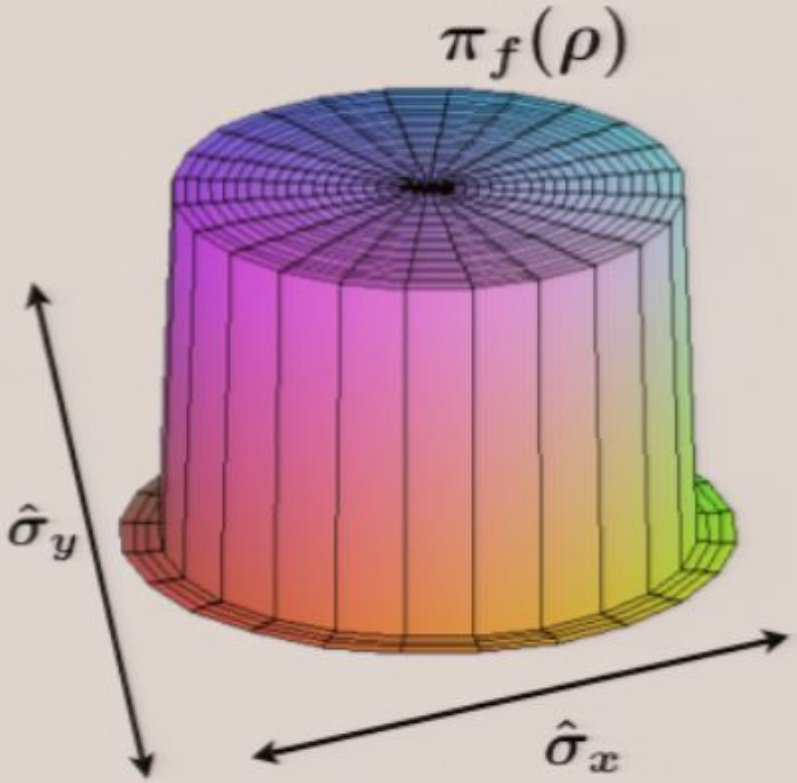
POVM operators are:

$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$$



Example: spin-½

Measure Y, X (ignore Z)

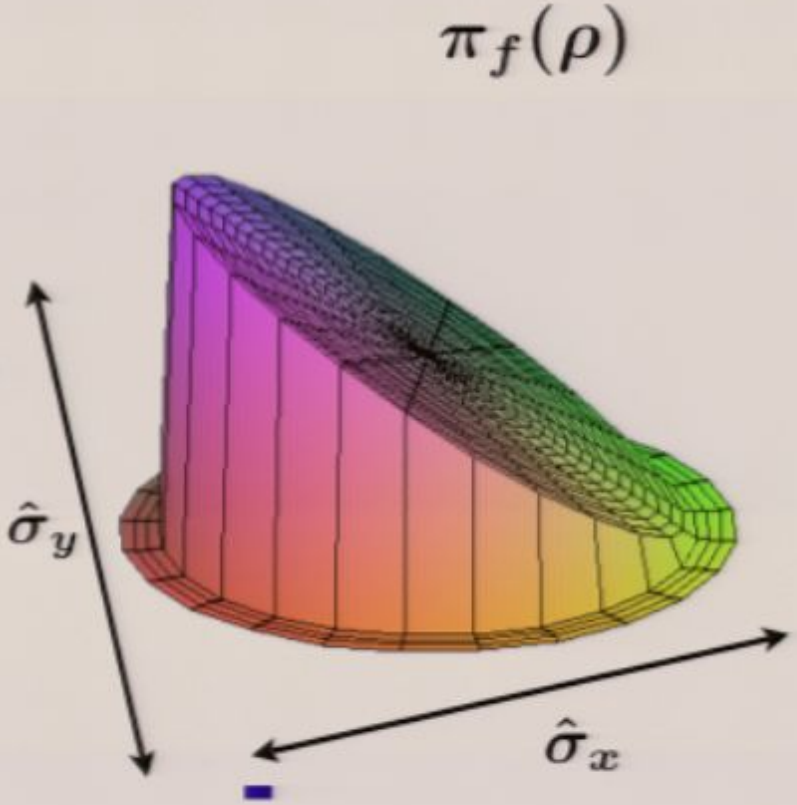
POVM operators are:

$$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$$



Example: spin-½

Measure Y, X (ignore Z)

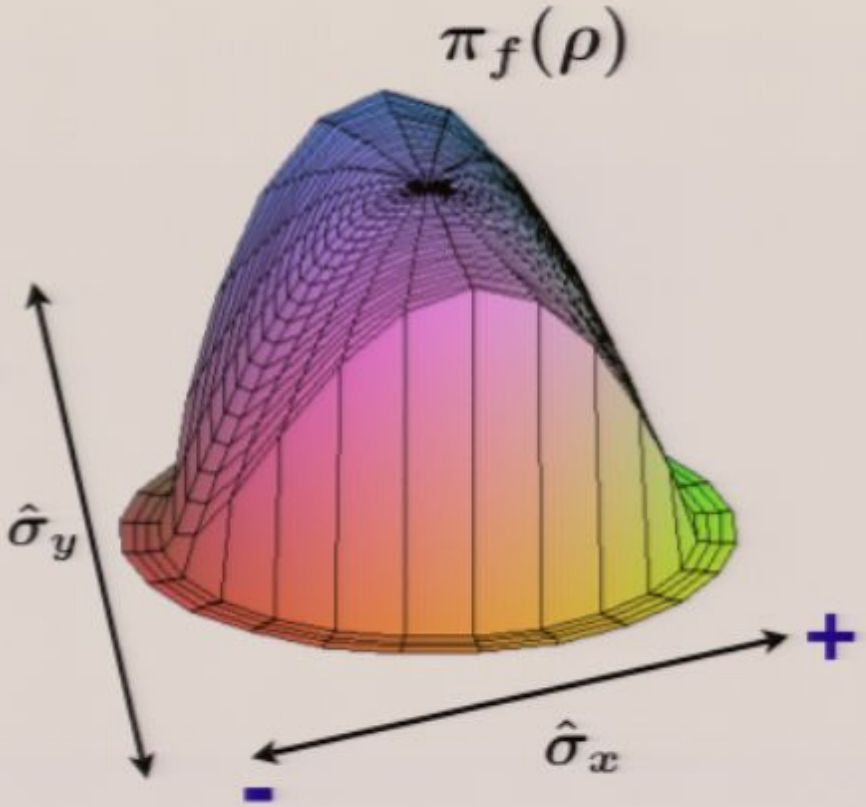
POVM operators are:

$$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$$



Example: spin-½

Measure Y, X (ignore Z)

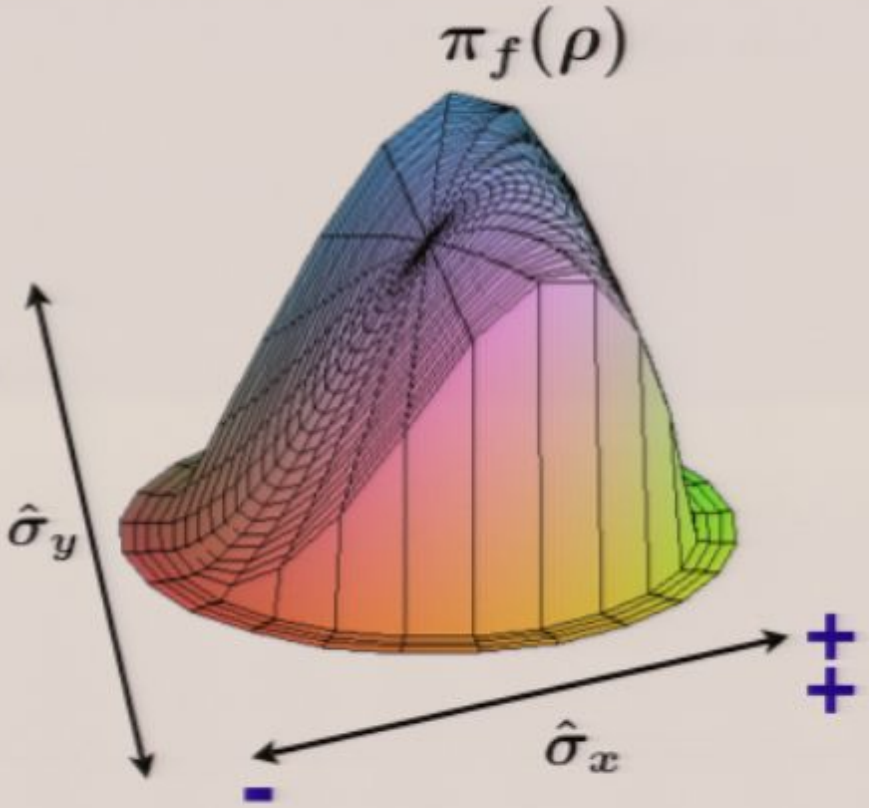
POVM operators are:

$$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$$



Example: spin-½

Measure Y, X (ignore Z)

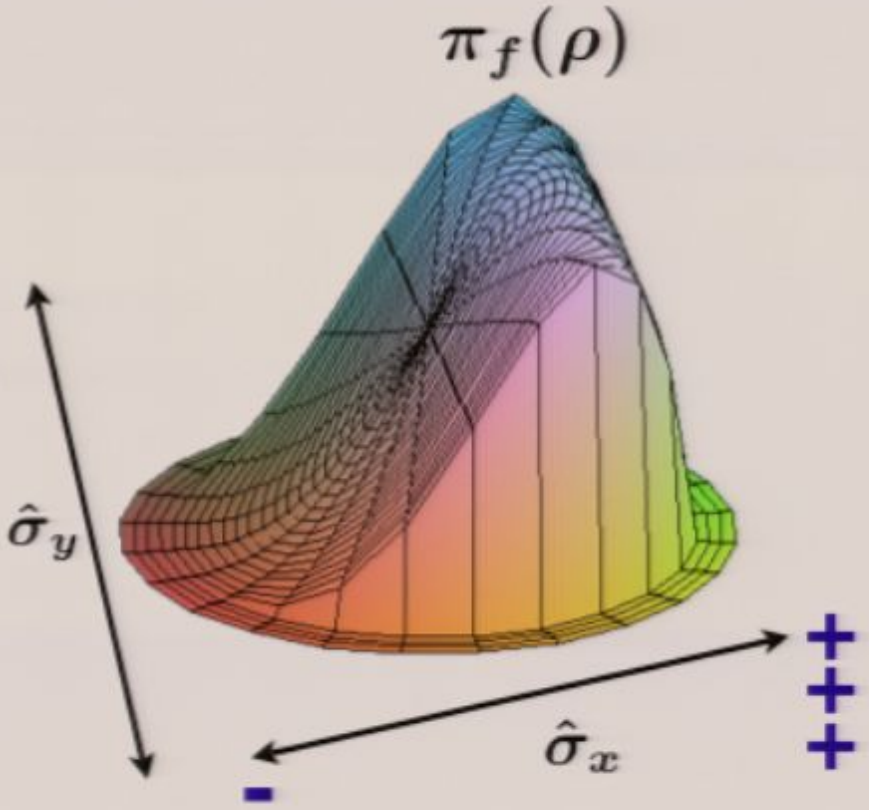
POVM operators are:

$$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$



Example: spin-½

Measure Y, X (ignore Z)

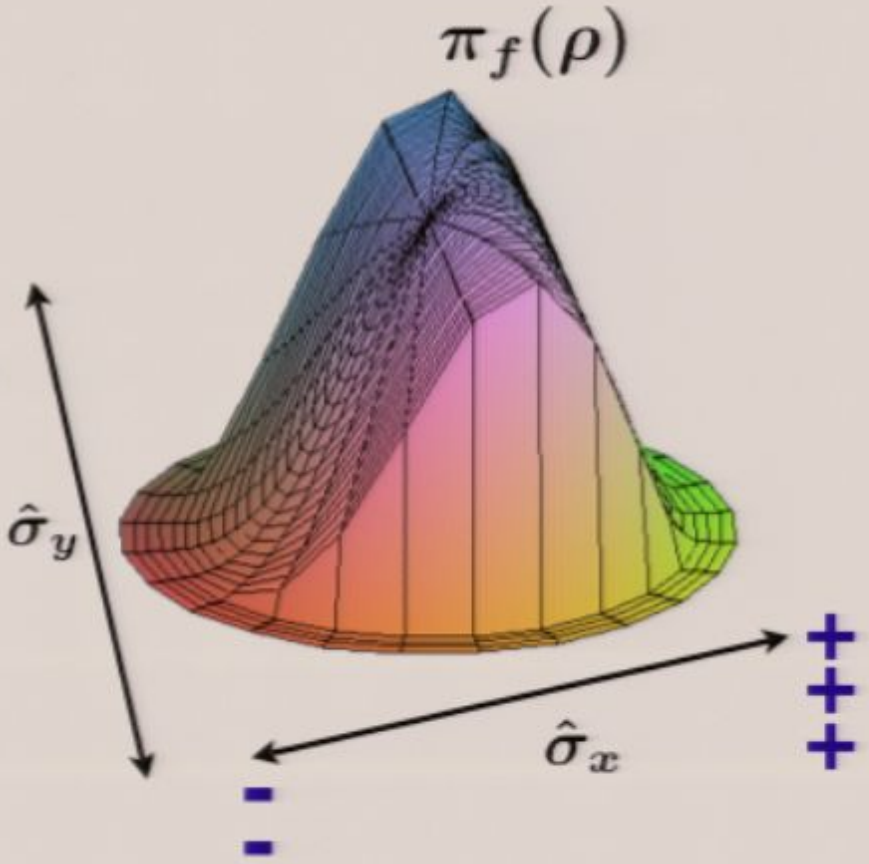
POVM operators are:

$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$$



Example: spin-½

Measure Y, X (ignore Z)

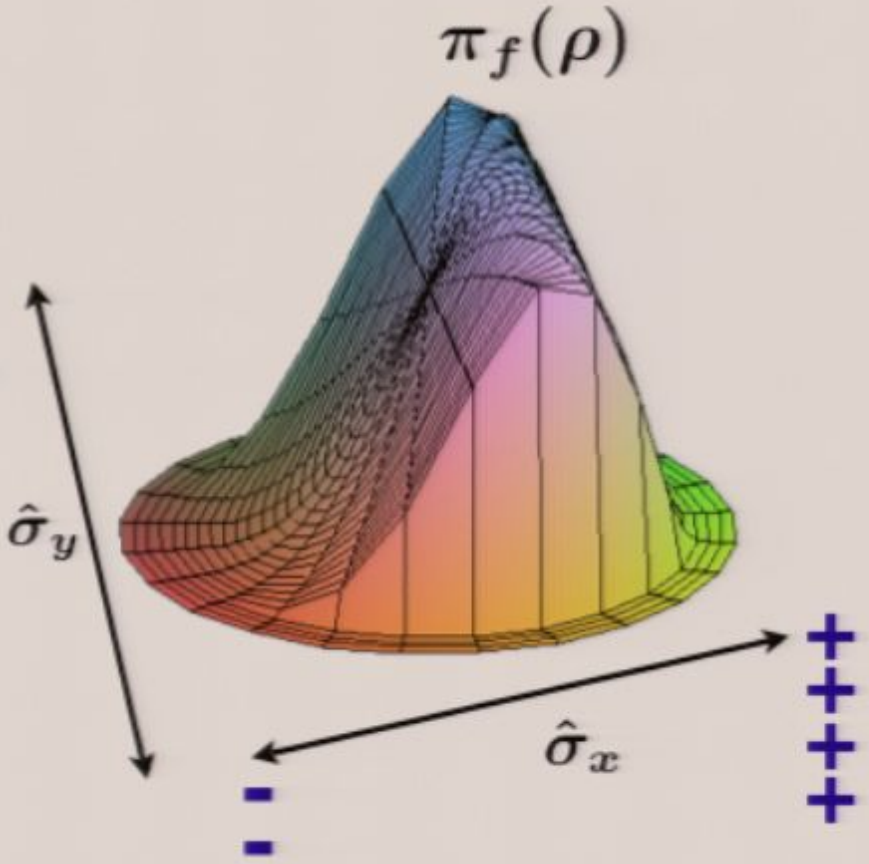
POVM operators are:

$$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$$



Example: spin-½

Measure Y, X (ignore Z)

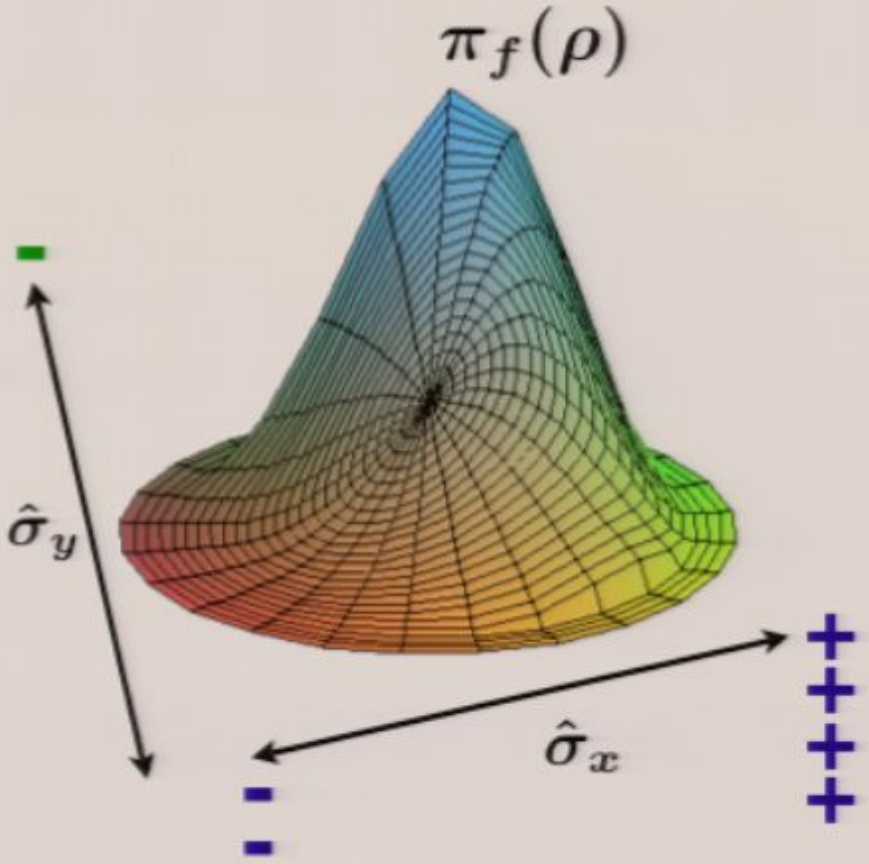
POVM operators are:

$$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$$



Example: spin-½

Measure Y, X (ignore Z)

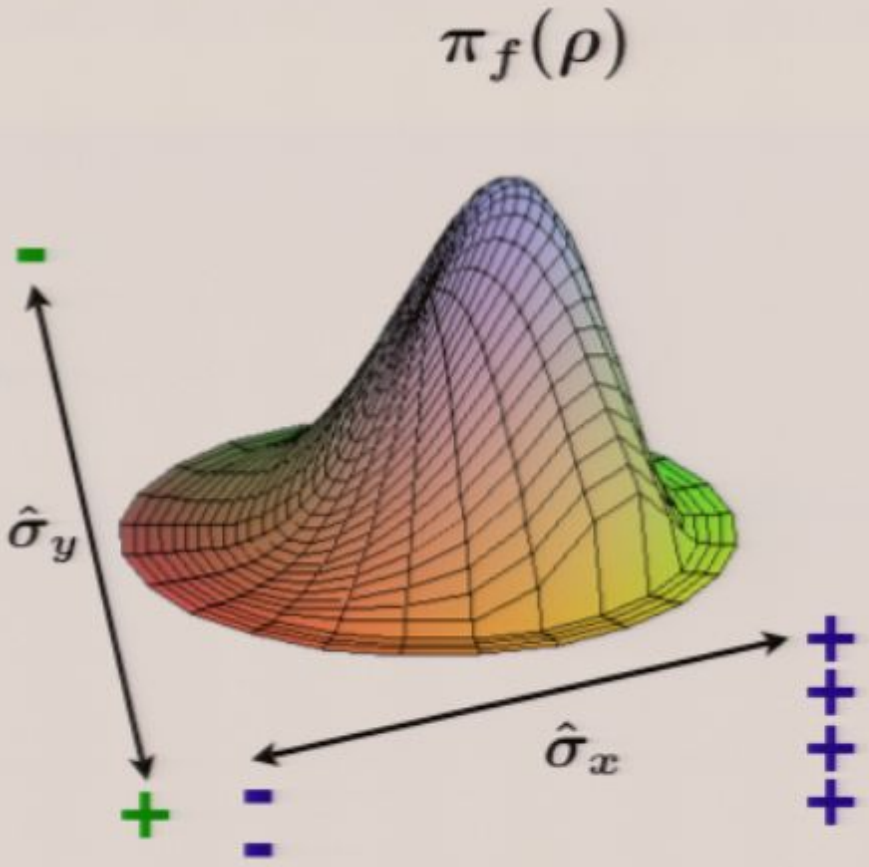
POVM operators are:

$$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$



Example: spin-½

Measure Y, X (ignore Z)

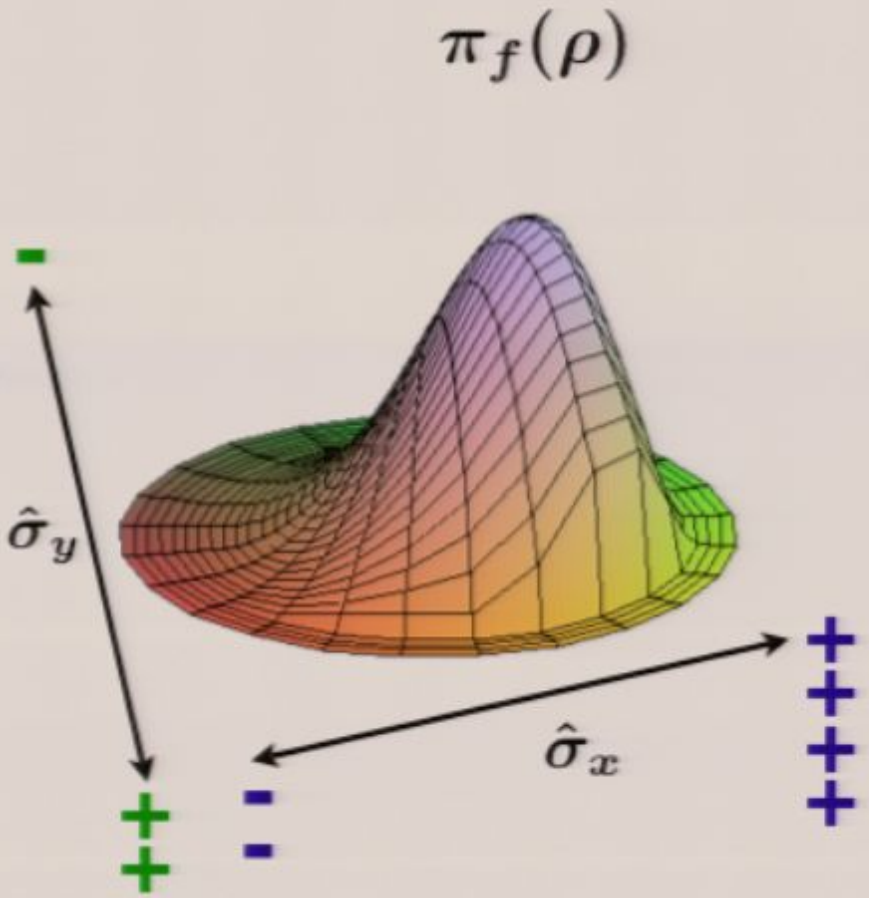
POVM operators are:

$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$$



Example: spin-½

Measure Y, X (ignore Z)

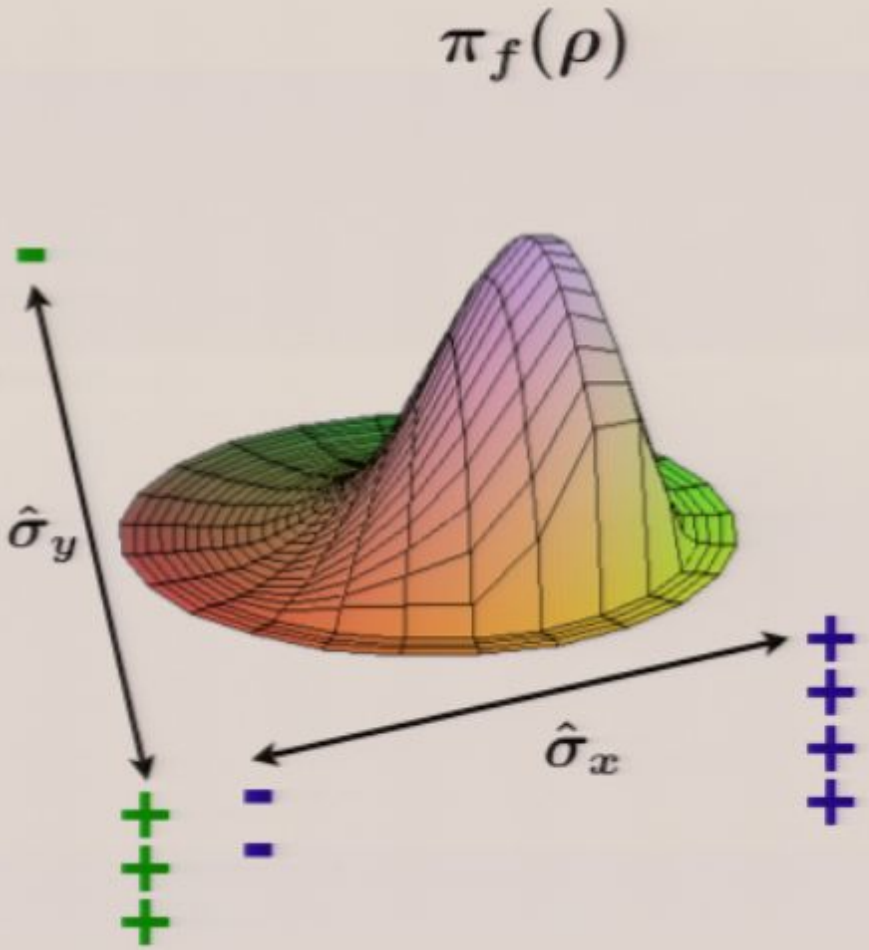
POVM operators are:

$$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$$



Example: spin-½

Measure Y, X (ignore Z)

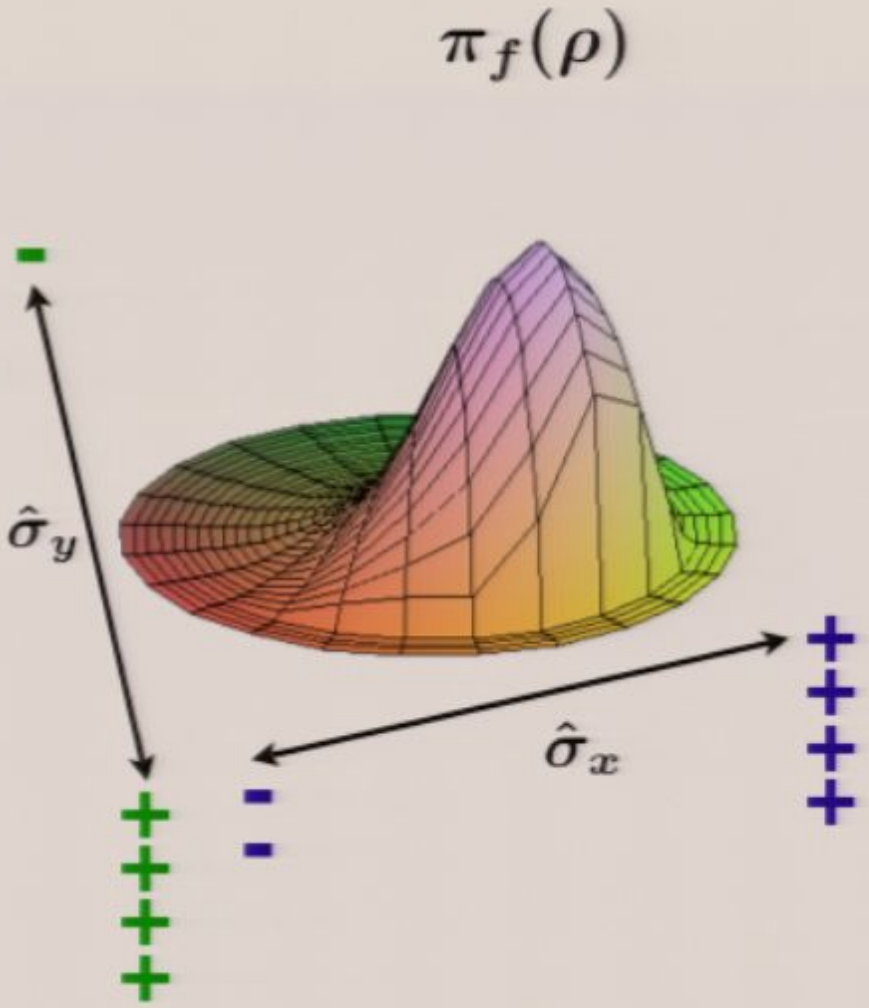
POVM operators are:

$$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$



Example: spin-½

Measure Y, X (ignore Z)

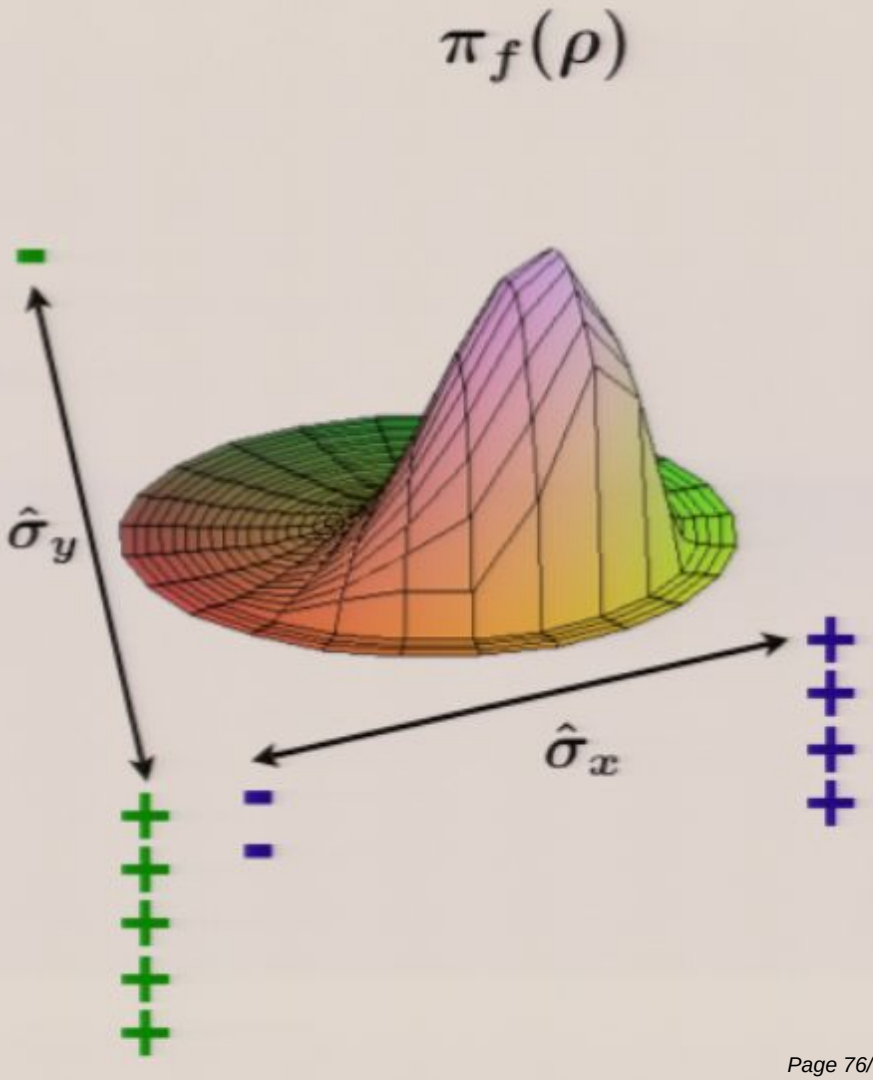
POVM operators are:

$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$$



Example: spin-½

Measure Y, X (ignore Z)

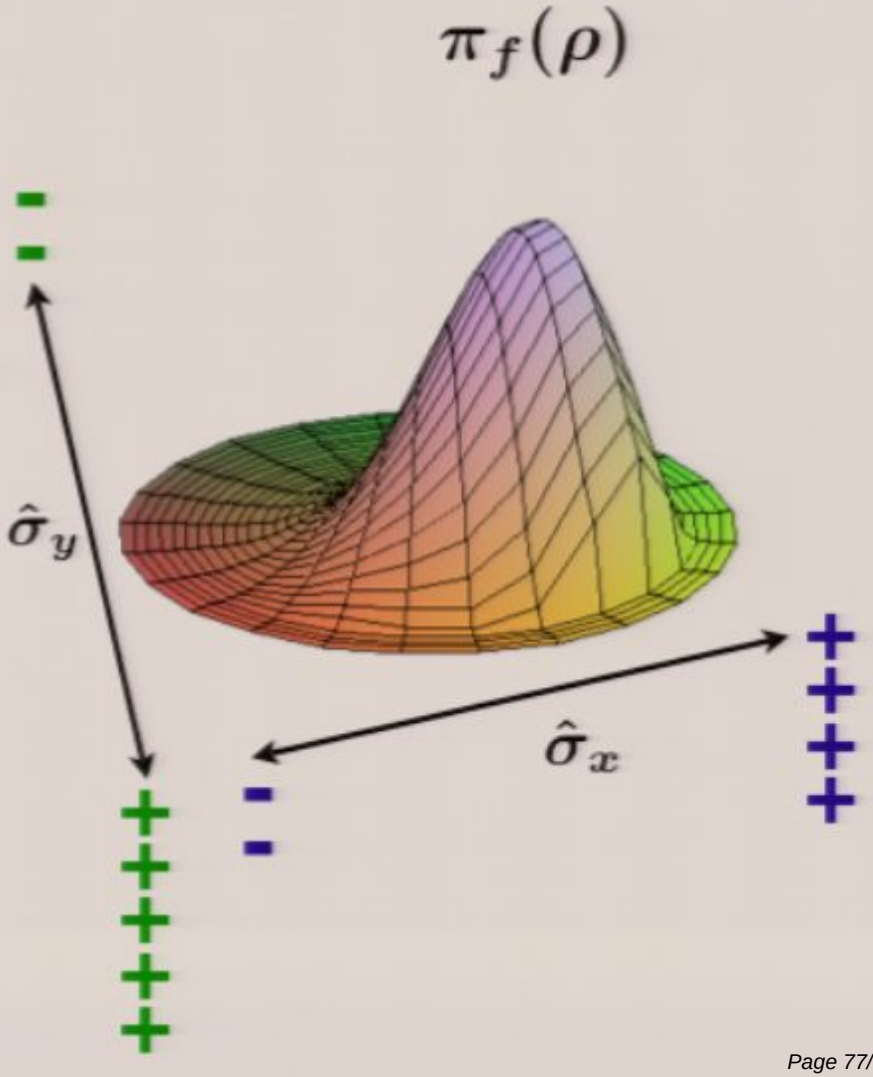
POVM operators are:

$$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$$



Example: spin-½

Measure Y, X (ignore Z)

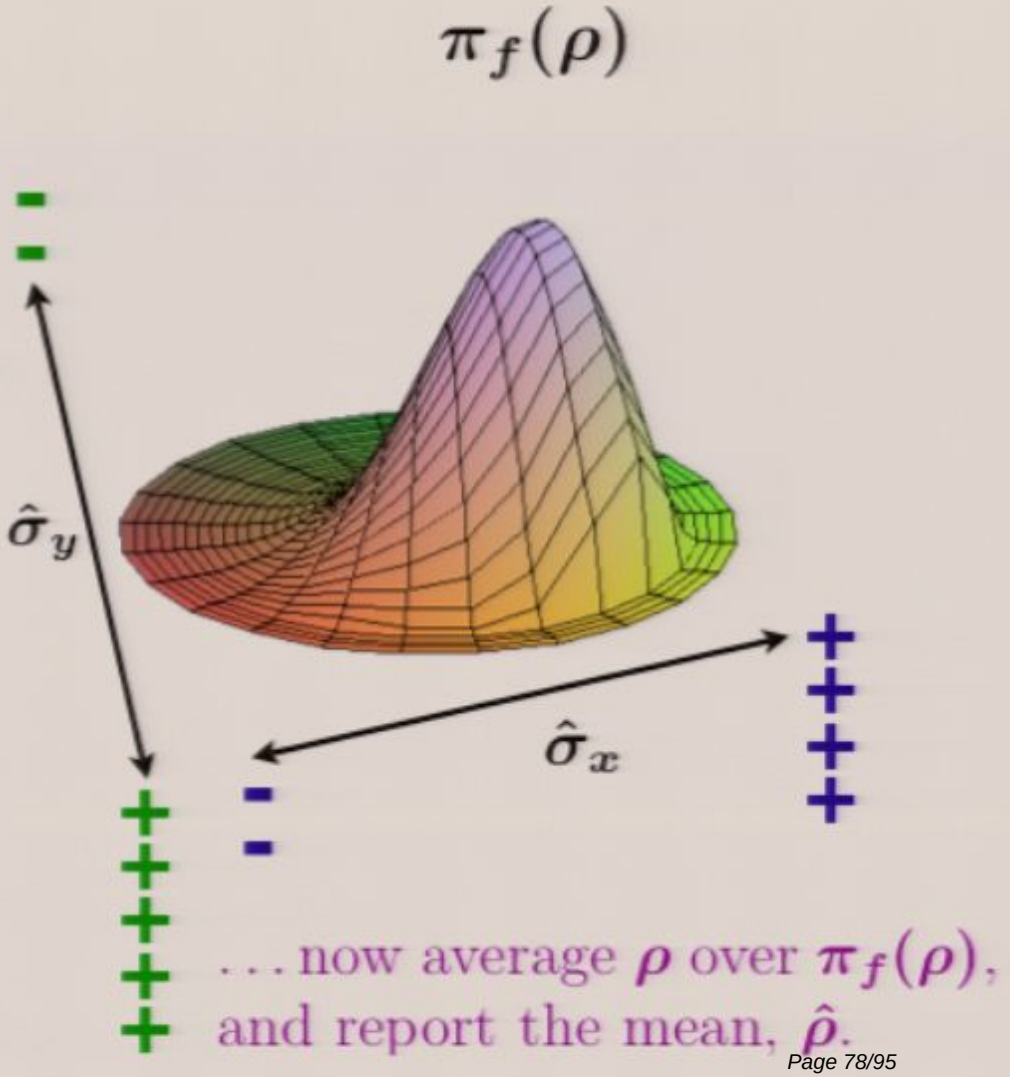
POVM operators are:

$$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$$

Bayesian Inference

1. Begin with a prior: $\pi_0(\rho)\mathrm{d}\rho$.
2. Make N measurements
   $\Rightarrow$ record: $\mathrm{M} = \left\{\hat{M}_1, \hat{M}_2, \ldots, \hat{M}_N\right\}$
3. Bayes' Rule: $\pi_f(\rho)\mathrm{d}\rho \propto \mathcal{L}(\rho)\pi_0(\rho)\mathrm{d}\rho$

$= \mathrm{Tr}(\hat{M}_1\rho)\mathrm{Tr}(\hat{M}_2\rho)\ldots\mathrm{Tr}(\hat{M}_N\rho)\pi_0(\rho)\mathrm{d}\rho$



Example: spin-½

Measure Y, X (ignore Z)

POVM operators are:

$\hat{E}_i \in \{|+\rangle\langle+|, |-\rangle\langle-|, |+i\rangle\langle+i|, |-i\rangle\langle-i|\}$

ρ_{BME} is strictly positive

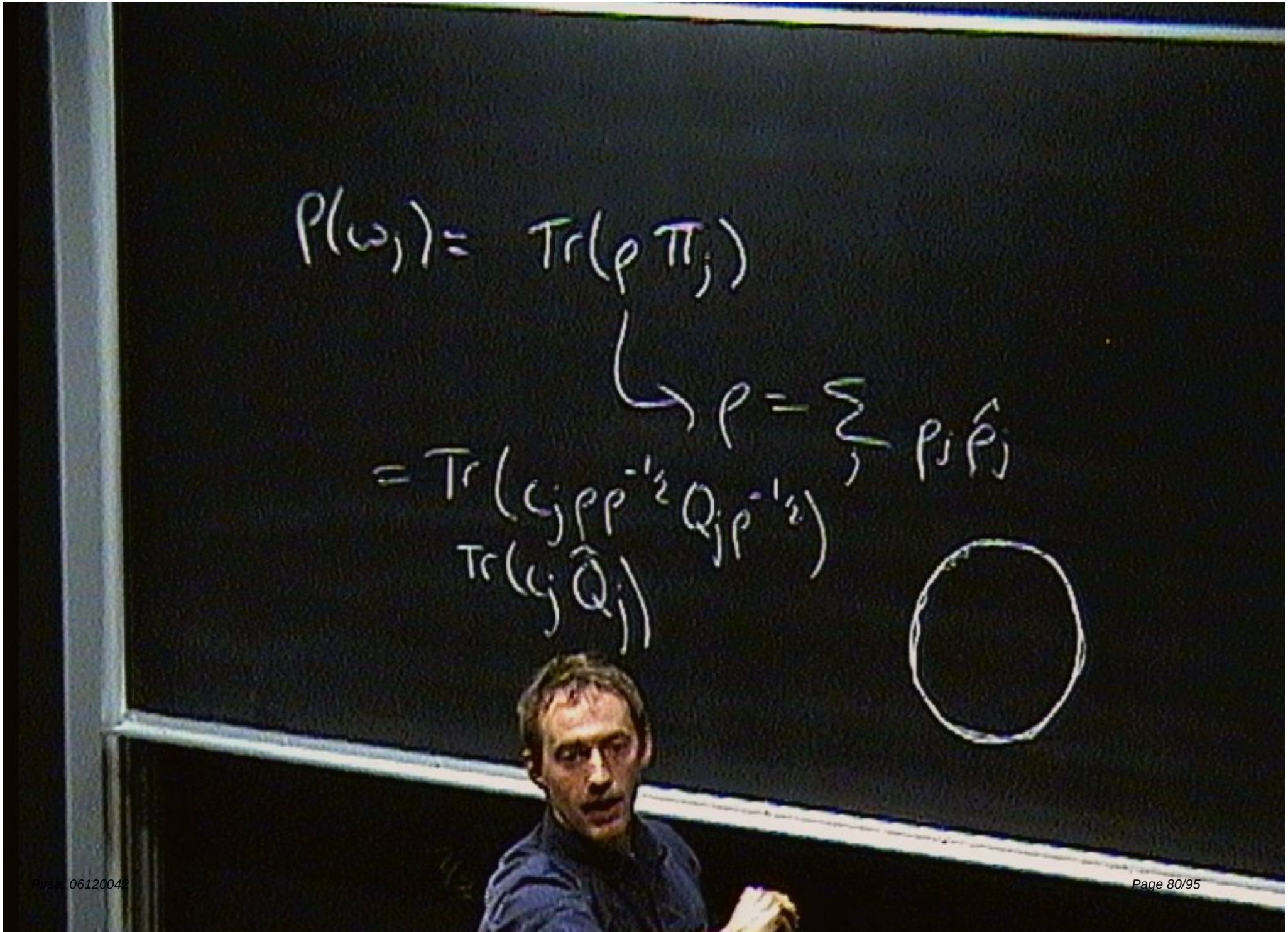
- BME never yields a rank-deficient estimate!
- Proof sketch:

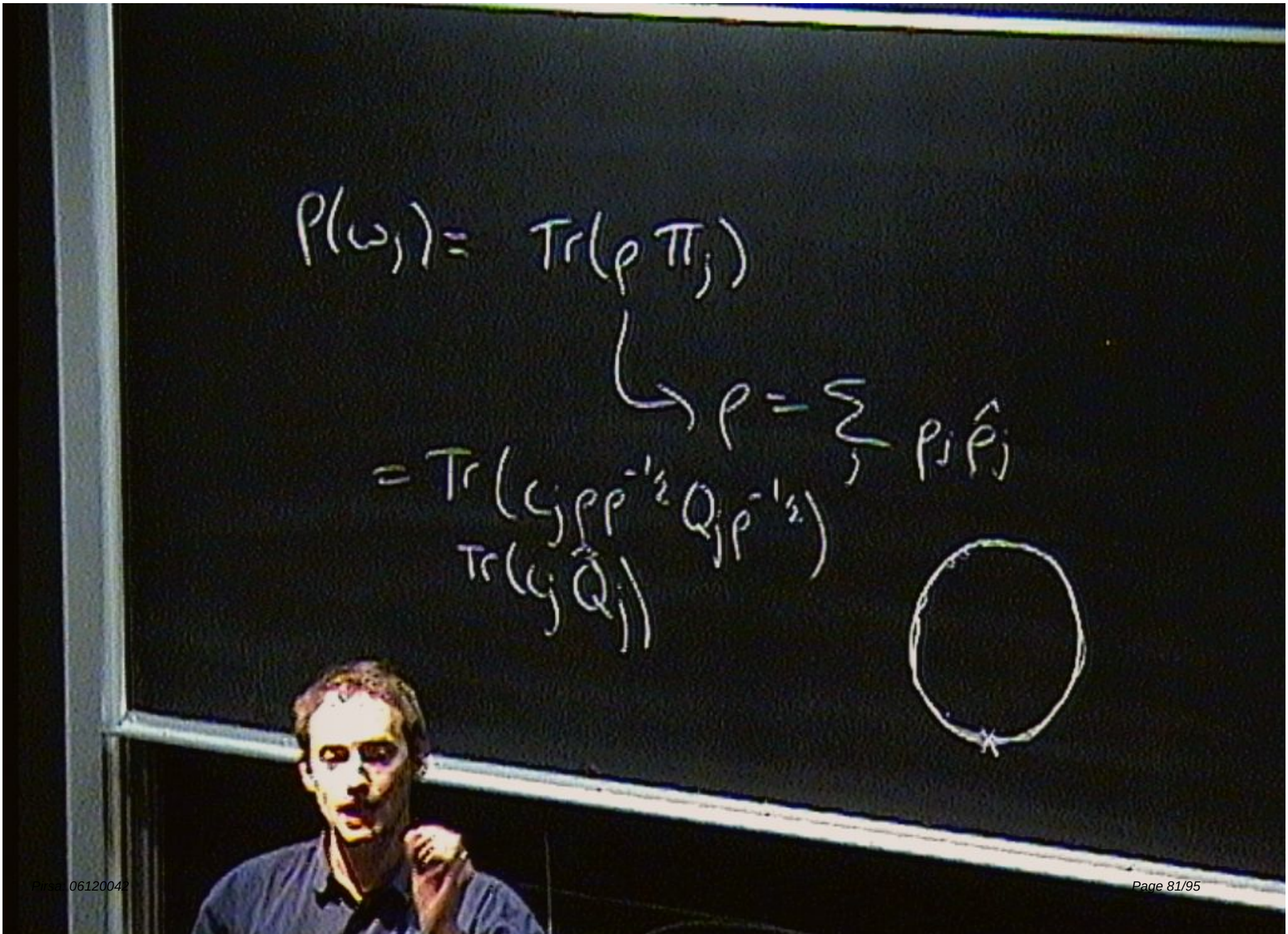
  For any $|\psi\rangle$, let $\boldsymbol{\pi_0(\rho)d\rho}$ have support on a measurable set of states *not* orthogonal to $|\psi\rangle$.

  Each measurement $\hat{\boldsymbol{M}}_{\boldsymbol{i}}$ removes a measure-zero set of states from the support.

  So $\boldsymbol{\pi_0(\rho)d\rho}$ contains states not orthogonal to $|\psi\rangle$.

  These contribute to $\boldsymbol{\rho_{\mathrm{BME}}}$, so $\langle\psi|\,\boldsymbol{\rho_{\mathrm{BME}}}\,|\psi\rangle > \mathbf{0}$.


$p(\omega_j) = \mathrm{Tr}(\rho \Pi_j)$
$\rho = \sum_j p_j \hat{\rho}_j$
$= \mathrm{Tr}(\omega_j \rho \rho^{-1/2} Q_j \rho^{-1/2})$
$\mathrm{Tr}(\omega_j \hat{Q}_j)$


$p(\omega_j) = \mathrm{Tr}(\rho \Pi_j)$
$\hookrightarrow \rho = \sum_j p_j \hat{\rho}_j$
$= \mathrm{Tr}(\omega_j \rho \rho^{-1/2} Q_j \rho^{-1/2})$
$\mathrm{Tr}(\omega_j \hat{Q}_j)$

ρ_{BME} is strictly positive

- BME never yields a rank-deficient estimate!
- Proof sketch:

  For any $|\psi\rangle$, let $\boldsymbol{\pi_0(\rho)\mathrm{d}\rho}$ have support on a measurable set of states *not* orthogonal to $|\psi\rangle$.

  Each measurement $\hat{M}_i$ removes a measure-zero set of states from the support.

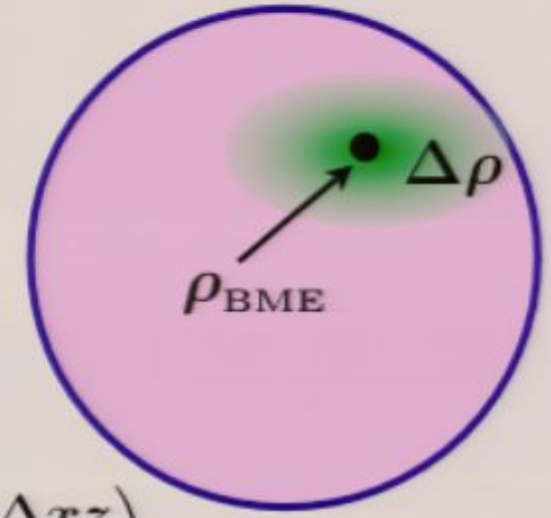
  So $\boldsymbol{\pi_0(\rho)\mathrm{d}\rho}$ contains states not orthogonal to $|\psi\rangle$.

  These contribute to $\boldsymbol{\rho_{\mathrm{BME}}}$, so $\langle\psi|\,\boldsymbol{\rho_{\mathrm{BME}}}\,|\psi\rangle > \mathbf{0}$.

Free error bars with every estimate!

- Estimate is a vector in Hilbert-Schmidt space:

$$\rho = \begin{pmatrix} \langle x \rangle \\ \langle y \rangle \\ \langle z \rangle \end{pmatrix}$$



- Uncertainty is a covariance matrix on H-S space:

$$\Delta\rho = \begin{pmatrix} \Delta x^2 & \Delta xy & \Delta xz \\ \Delta xy & \Delta y^2 & \Delta yz \\ \Delta xz & \Delta yz & \Delta z^2 \end{pmatrix}$$

- I.e., a *superoperator*, with Jamiolkowski form:

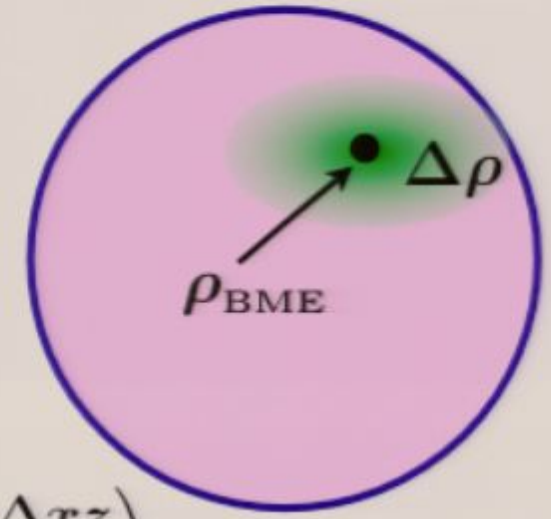
$$\Delta\rho_{\mathrm{BME}} = \overline{\rho \otimes \rho} - \overline{\rho} \otimes \overline{\rho}$$

$$= \int (\rho \otimes \rho)\pi(\rho)\mathrm{d}\rho - \rho_{\mathrm{BME}} \otimes \rho_{\mathrm{BME}}$$

Free error bars with every estimate!

- Estimate is a vector in Hilbert-Schmidt space:

$$\rho = \begin{pmatrix} \langle x \rangle \\ \langle y \rangle \\ \langle z \rangle \end{pmatrix}$$



- Uncertainty is a covariance matrix on H-S space:

$$\Delta\rho = \begin{pmatrix} \Delta x^2 & \Delta xy & \Delta xz \\ \Delta xy & \Delta y^2 & \Delta yz \\ \Delta xz & \Delta yz & \Delta z^2 \end{pmatrix}$$

- I.e., a *superoperator*, with Jamiolkowski form:

$$\Delta\rho_{\mathrm{BME}} = \overline{\rho \otimes \rho} - \overline{\rho} \otimes \overline{\rho}$$

$$= \int (\rho \otimes \rho)\pi(\rho)\mathrm{d}\rho - \rho_{\mathrm{BME}} \otimes \rho_{\mathrm{BME}}$$

Accuracy:

BME optimizes every operational divergence

If you *know* ρ, the optimal estimate is $\sigma = \rho$.

But what if your knowledge is uncertain (probabilistic)?

$\Longrightarrow$ the state could be ρ_i with probability π_i.

KEY FACT: $f(\rho : \sigma) = \mathrm{Tr}[\rho \mathcal{R}(\sigma)]$ is *linear* in ρ.

$\Longrightarrow$ *expected* utility is

$$\overline{f} = \sum \pi_i f(\rho_i : \sigma) = f\left(\sum \pi_i \rho_i : \sigma\right)$$

$\Longrightarrow$ optimal estimate is $\sigma = \overline{\rho} \equiv \sum \pi_i \rho_i$

This is fairly straightforward, but tedious... see quant-ph/0603116

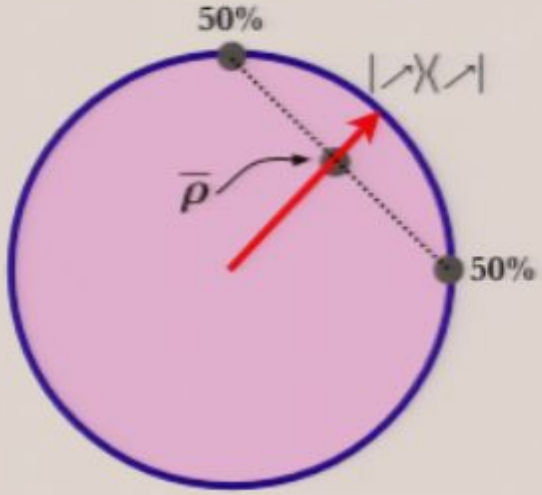
If unknown ρ was selected from distribution $\pi_0(\rho)d\rho$, **and** measurements $\mathcal{M} = \{E_1, E_2 \ldots\}$ were made, **then**:

1. Your knowledge is $\pi(\rho)d\rho = \frac{p(\mathcal{M}|\rho)\pi_0(\rho)d\rho}{\int p(\mathcal{M}|\rho)\pi_0(\rho)d\rho}$.
2. The optimal estimate is $\overline{\rho} = \int \rho\pi(\rho)d\rho$.

But isn't the mean an obvious choice?

But isn't the mean an obvious choice? *NO*

(a) Suppose we optimize fidelity.

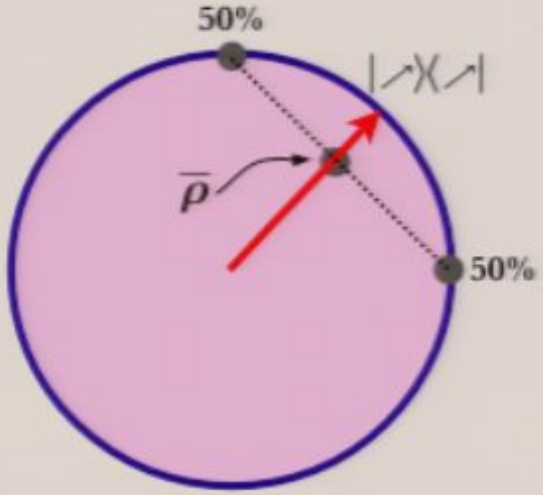


$$\overline{F}\left(\rho_i, \overline{\rho}\right) = \tfrac{3}{4}$$
$$\overline{F}\left(\rho_i, |\nearrow\rangle\langle\nearrow|\right) = \tfrac{3+2\sqrt{2}}{4+2\sqrt{2}} \approx 0.85$$

MORAL: fidelity measures how well σ *simulates* ρ, not how well it *estimates* ρ.

But isn't the mean an obvious choice? *NO*

(a) Suppose we optimize fidelity.



$$\overline{F}\left(\rho_i, \overline{\rho}\right) = \tfrac{3}{4}$$
$$\overline{F}\left(\rho_i, |\nearrow\rangle\langle\nearrow|\right) = \tfrac{3+2\sqrt{2}}{4+2\sqrt{2}} \approx 0.85$$

MORAL: fidelity measures how well $\boldsymbol{\sigma}$ *simulates* $\boldsymbol{\rho}$, not how well it *estimates* $\boldsymbol{\rho}$.

(b) Suppose we assume the future will look (statistically) just like the past.

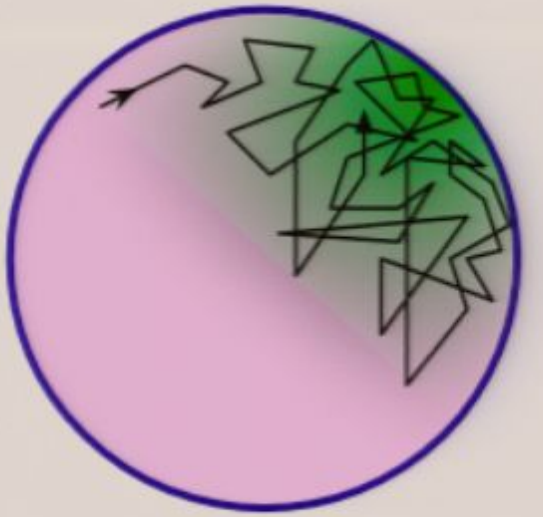
- We know what future datasets will look like
- We can add them to measurement record
- $\mathcal{M} \longrightarrow \mathcal{M} \cup \mathcal{M} \cup \mathcal{M} \cup \ldots$
- $p(\mathcal{M}|\rho) \longrightarrow p(\mathcal{M}|\rho)^{\infty}$
- So $\pi(\rho)d\rho \rightarrow \delta\left(\rho - \rho_{\mathrm{MLE}}\right)$

MORAL: MLE can be derived by assuming this "frequentist axiom".

(This explains a lot...)

Implementing BME
via Metropolis-Hastings

- Crawl over Hilbert-Schmidt space, integrating as we go.
- Jump from $\rho \to \rho'$ w/prob $p_{\text{jump}} = \max\left(\frac{\mathcal{L}(\rho')}{\mathcal{L}(\rho)}, 1\right)$



- Seems to scale as N^4-N^6 (slower than MLE)

Is BME really optimal?

1. **Motivation**: problems with MLE.
2. **Foundation**: quantum scoring rules.
3. **Solution**: the BME algorithm (& its features)
4. **Testing**: dueling procedures! (numerics)

Is BME really optimal?

1. **Motivation**: problems with MLE.

2. **Foundation**: quantum scoring rules.

3. **Solution**: the BME algorithm (& its features)

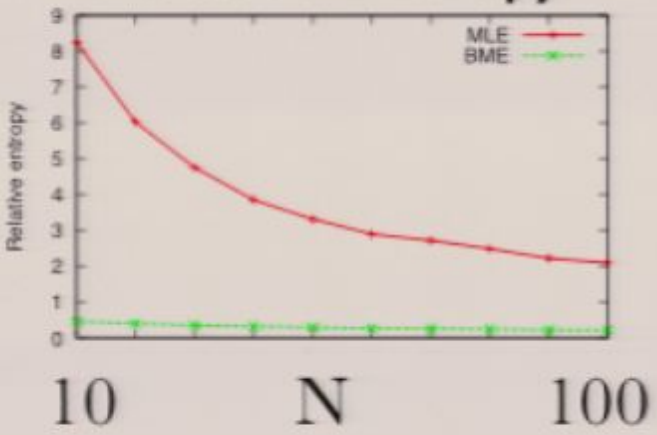
4. **Testing**: dueling procedures! (numerics)

How good is MLE?

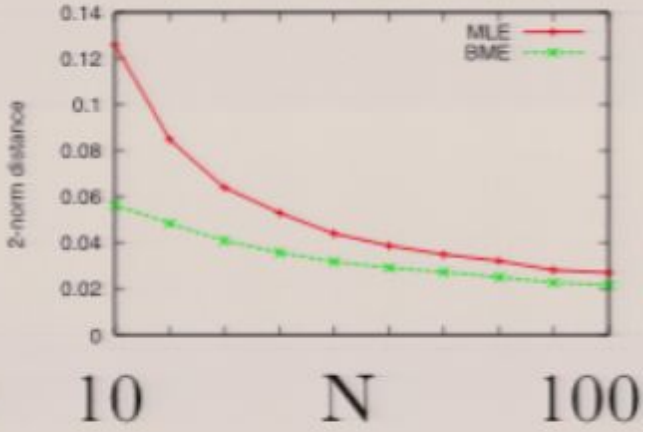
NUMERICS:

1. Generate random (Hilbert-Schmidt measure) mixed states $\boldsymbol{\rho}$ for $\boldsymbol{n}$ qubits,
2. Measure each of 4^{n-1} Pauli observables $\boldsymbol{N}$ times,
3. Analyze measurement record to get $\boldsymbol{\sigma}$,
4. Compare $\boldsymbol{\sigma}$ to $\boldsymbol{\rho}$ using:
   (a) L2-norm, $\mathbf{Tr}\,[(\rho-\sigma)^2]$,
   (b) relative entropy, $\mathbf{Tr}\,[\rho(\ln\rho-\ln\sigma)]$,
   (c) fidelity, $\mathbf{Tr}\left[\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right]^2$
   (d) L1-norm, $\mathbf{Tr}\,[|\rho-\sigma|]$.

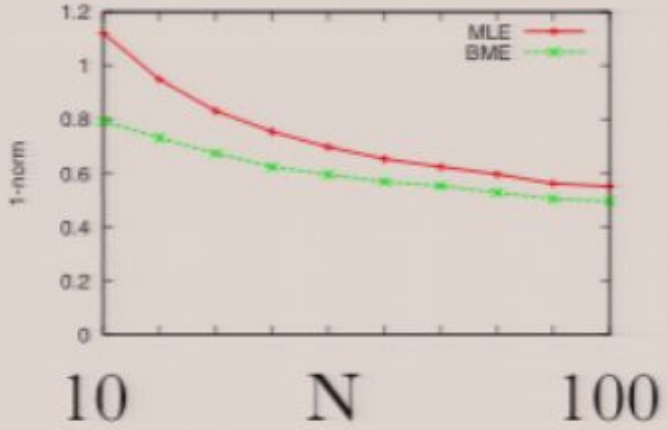
Relative Entropy



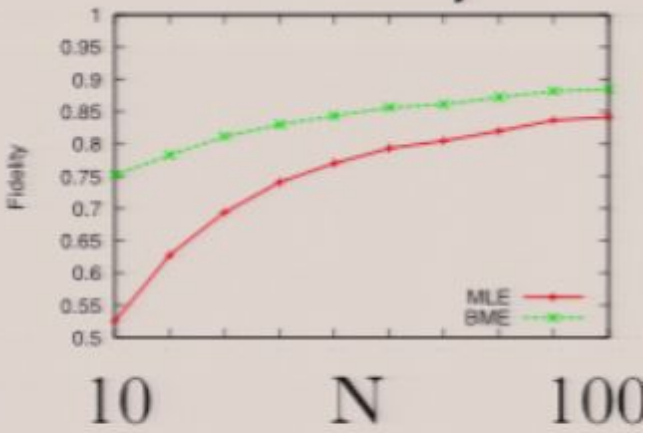
L-2 distance



Trace Distance



Fidelity

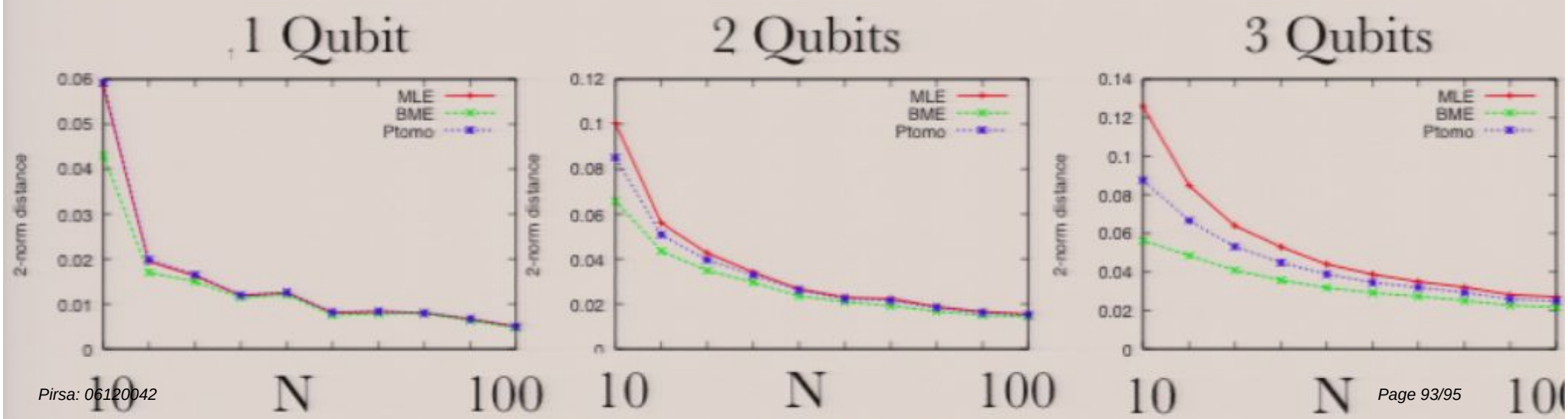


All data is for 3 qubits

Quick & Dirty Tomography

1. MLE gets computationally difficult for large systems.
2. BME is even harder!
3. $\frac{1}{\sqrt{3}}\left(|\text{D.James}\rangle + |\text{T.Havel}\rangle + |\text{P.Jessen}\rangle\right)$ suggested "quick 'n' dirty tomography".

$$\sigma_{\text{tomo}} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & -\lambda_3 & \\ & & & -\lambda_4 \end{pmatrix} \longrightarrow \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & 0 & \\ & & & 0 \end{pmatrix} \longrightarrow \frac{1}{\lambda_1 + \lambda_2} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & 0 & \\ & & & 0 \end{pmatrix} = \sigma.$$



All data is L_2 distance

Conclusions

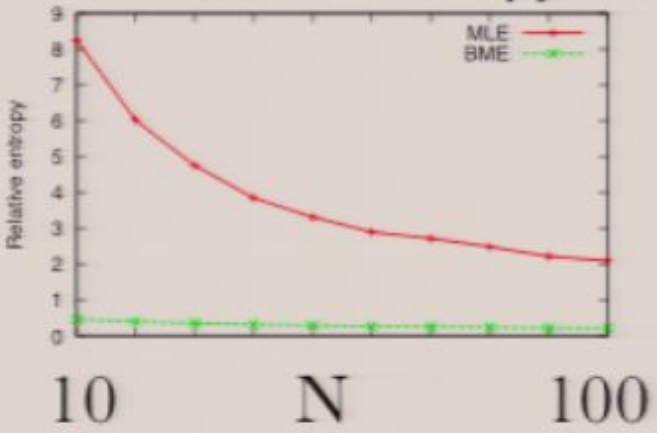
- It's worth thinking carefully and deeply about state estimation.
- Scoring rules provide a *reliable* experimental way to compare multiple estimates .
- BME is optimal, and a baseline for evaluating other (more efficient) approaches.
- Quick & dirty methods can outperform MLE.

How good is MLE?

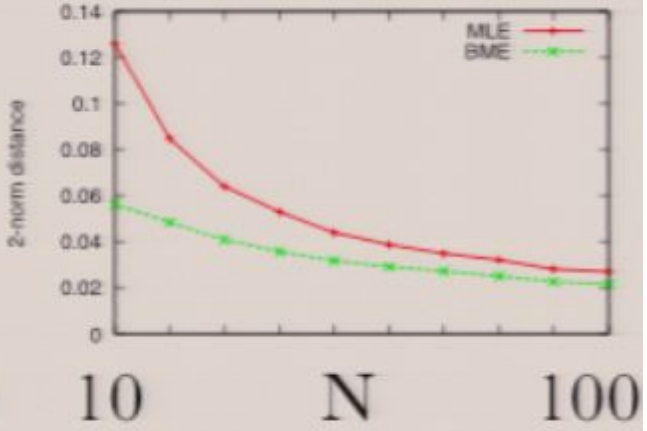
NUMERICS:

1. Generate random (Hilbert-Schmidt measure) mixed states $\boldsymbol{\rho}$ for $\boldsymbol{n}$ qubits,
2. Measure each of 4^{n-1} Pauli observables $\boldsymbol{N}$ times,
3. Analyze measurement record to get $\boldsymbol{\sigma}$,
4. Compare $\boldsymbol{\sigma}$ to $\boldsymbol{\rho}$ using:
   (a) L2-norm, $\mathbf{Tr}\,[(\rho-\sigma)^2]$,
   (b) relative entropy, $\mathbf{Tr}\,[\rho(\ln\rho-\ln\sigma)]$,
   (c) fidelity, $\mathbf{Tr}\left[\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right]^2$
   (d) L1-norm, $\mathbf{Tr}\,[|\rho-\sigma|]$.

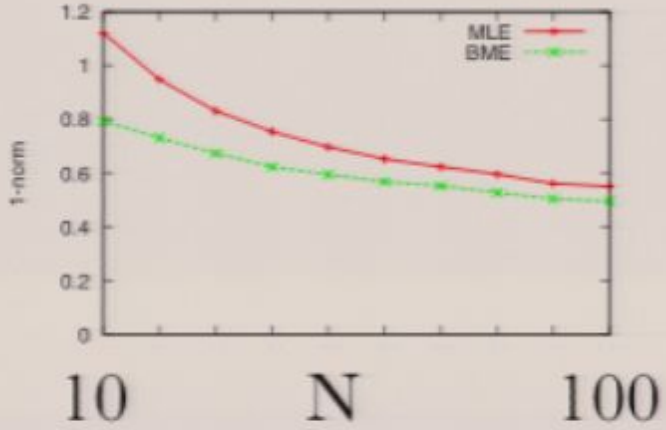
Relative Entropy



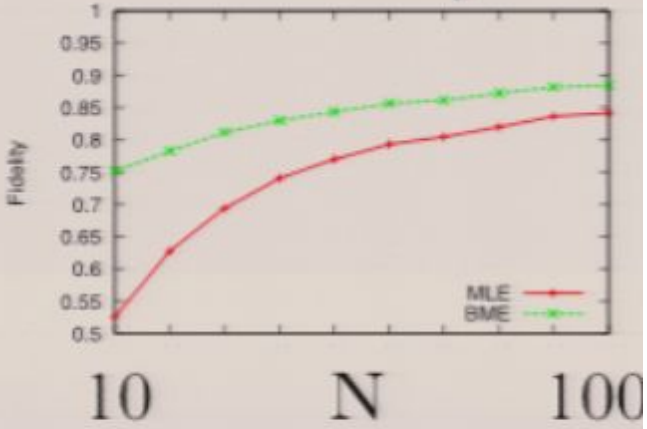
L-2 distance



Trace Distance



Fidelity



All data is for 3 qubits