SAND2015-10389PE

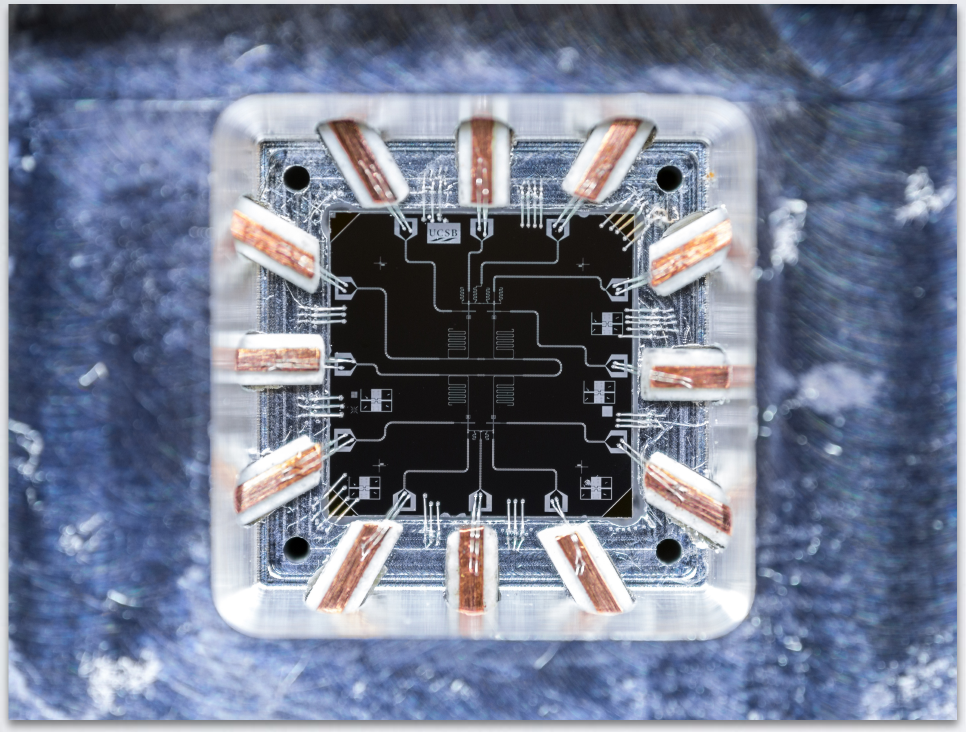
ON THE EDGE: STATE TOMOGRAPHY, BOUNDARIES, AND MODEL SELECTION

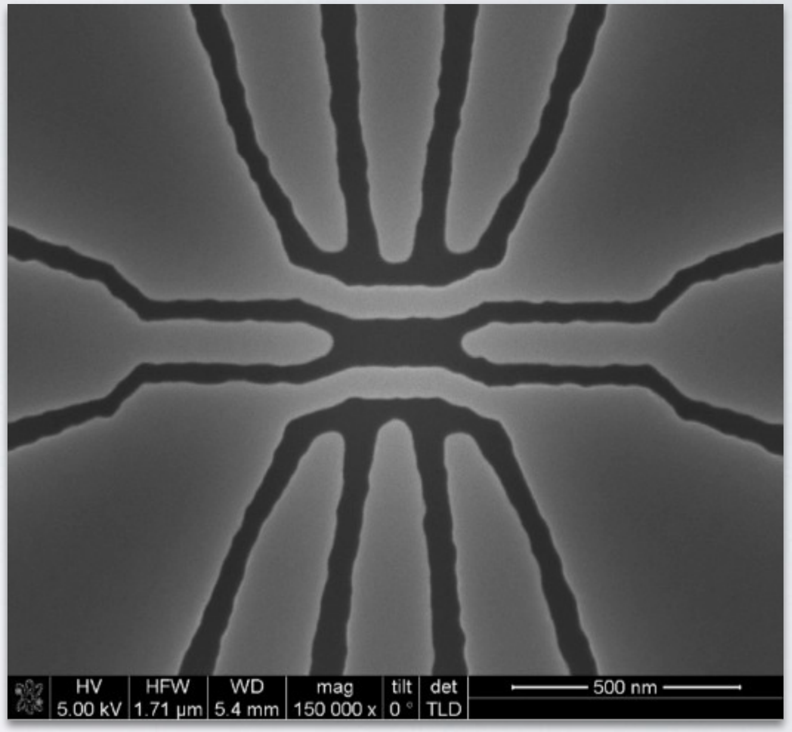
Travis L Scholten
Center for Computing Research
Sandia National Labs

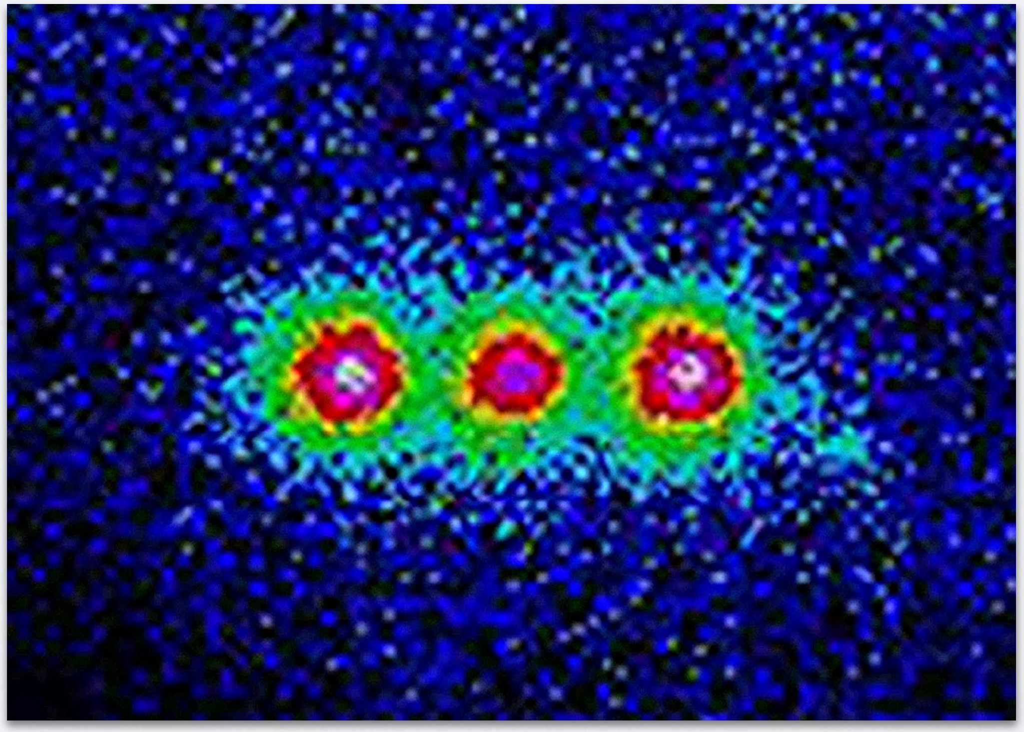
CQuIC Talk
University of New Mexico
2015 December 2

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

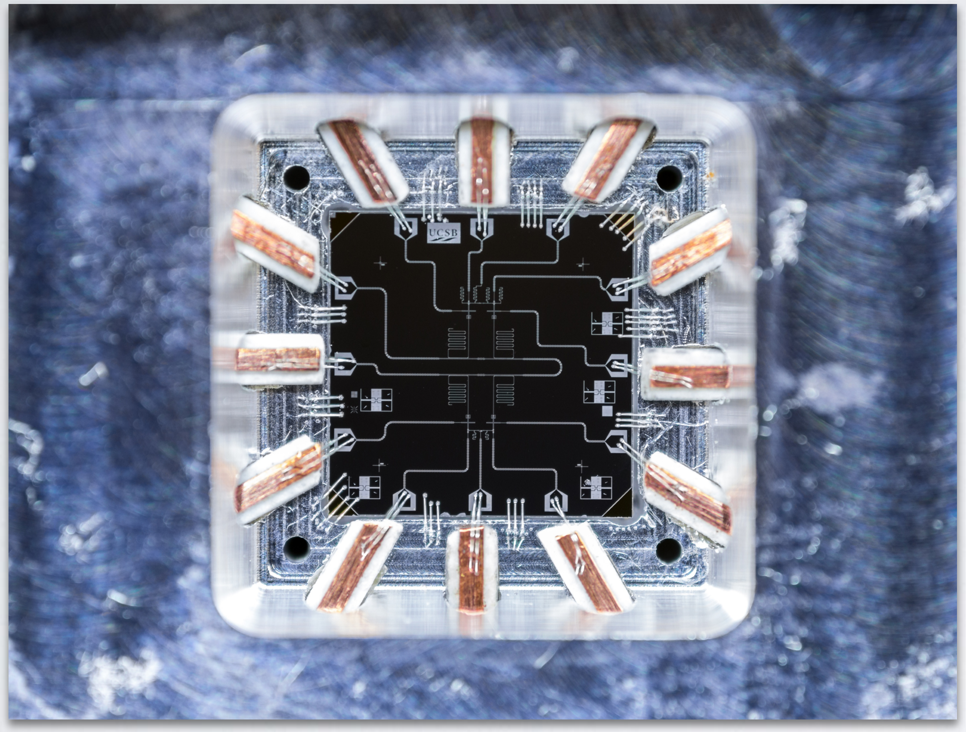
I am interested in tomography, the characterization of quantum systems.







Characterizing a system means estimating/inferring something about it.



Tomography is a *statistical inference* problem.

Data → Estimator → Estimates

(Experiment) (Inference) (Things with hats)

POVM $\{E_j\}$

$\hat{\rho}$

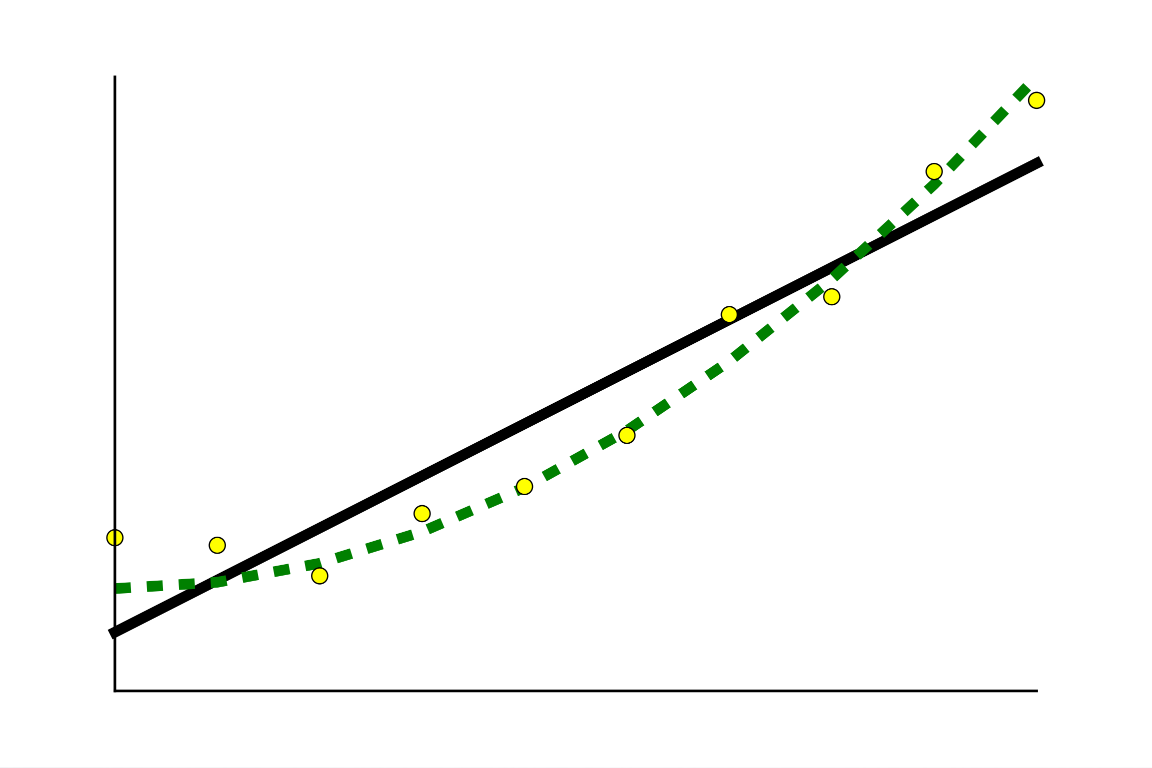
The “best” estimator is very accurate…

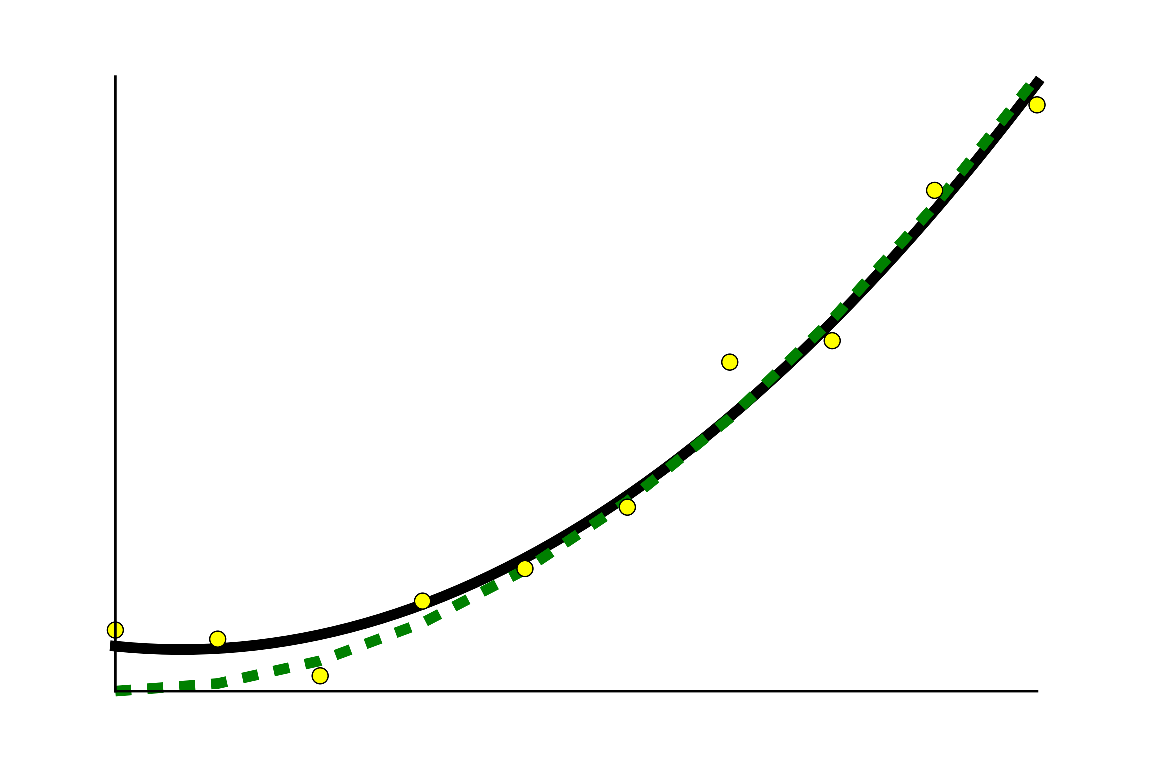
Many measures of accuracy:

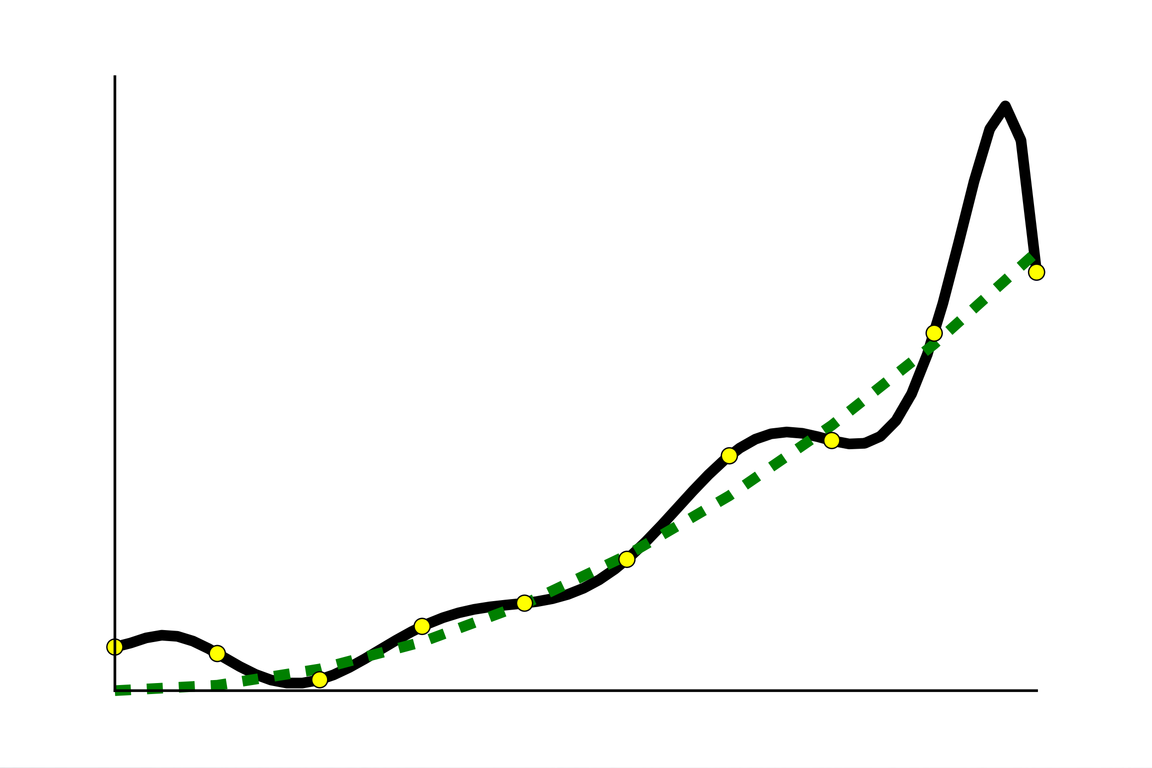
Quantum Fidelity	Relative Entropy
Trace Distance	Hilbert-Schmidt Distance

We seek high accuracy *relative to an* ***unknown*** *truth.*

The "ideal" estimator would be very accurate… and would not fit noise in the data.

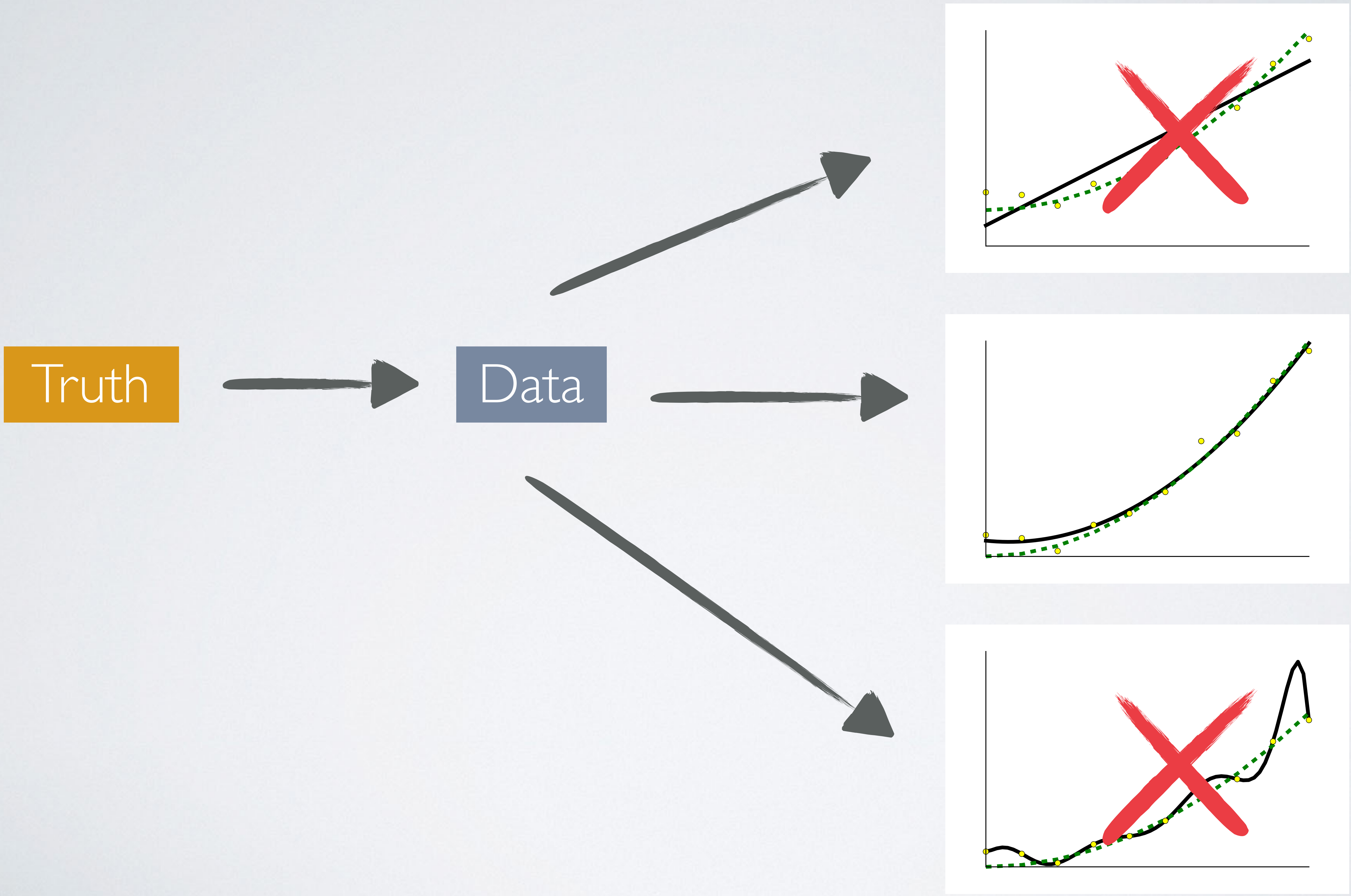






"Ideal" *impossible to achieve!*

We do not have truth, only data. How do we be accurate and fit well? Model selection.



People already use model selection in quantum information…but are they justified in doing so?

"Error models in quantum computation: an application of model selection" (Schwarz/van Enk, **2013**)

"Rank-based model selection for multiple ions quantum tomography" (Guta *et. al.*, **2012**)

"When quantum tomography goes wrong: drift of quantum sources and other errors" (van Enk/Blume-Kohout, **2013**)

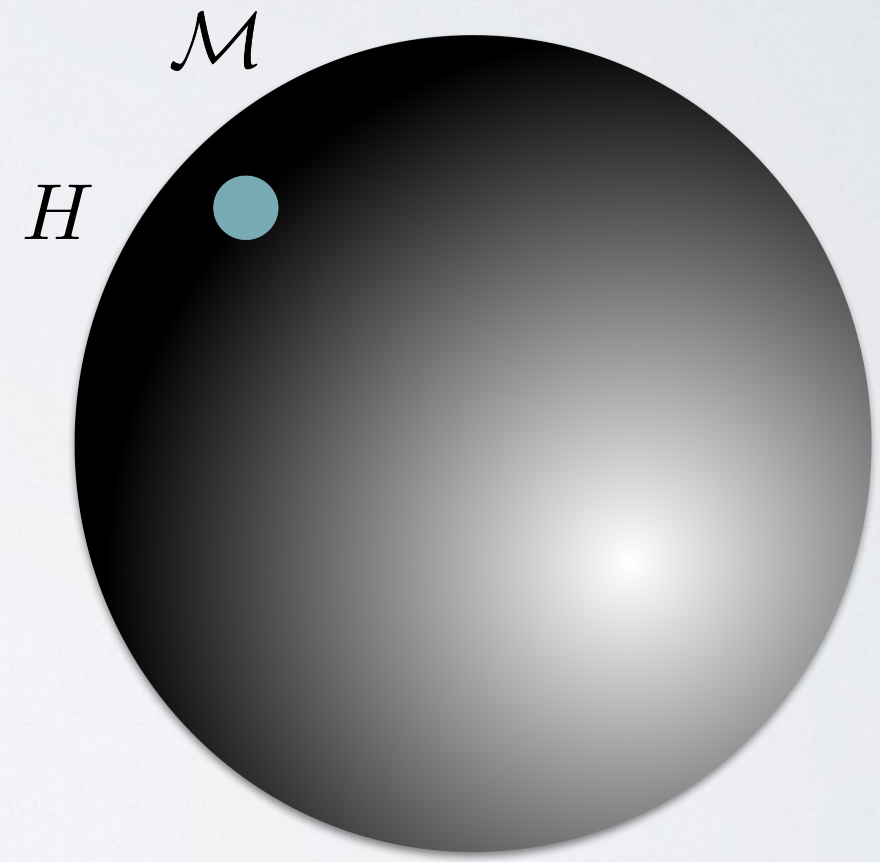
Model selection techniques currently used in quantum tomography **may have problems**.

Such as loglikelihood ratio tests, or Akaike's AIC.

Quantum information makes connections to statistical inference in many ways.

Model = parametrized family of probability distributions

Hypothesis = point in the model



Distributions via the Born rule:

$$Pr(E) = tr(\rho E)$$

Given some data, plausibility of models/hypotheses is quantified by their *likelihood*.

What is the probability assigned to the data seen?

Hypothesis: Just compute it! $$\mathcal{L}(H) = Pr(\mathrm{Data}|H)$$

Models: Just maximize! $$\mathcal{L}(\mathcal{M}) = \max_{H \in \mathcal{M}} Pr(\mathrm{Data}|H)$$

We use likelihoods to compare hypotheses/models and to make estimates.

Quantum information makes connections to statistical inference in many ways.

State discrimination

is an instance of

simple hypothesis testing

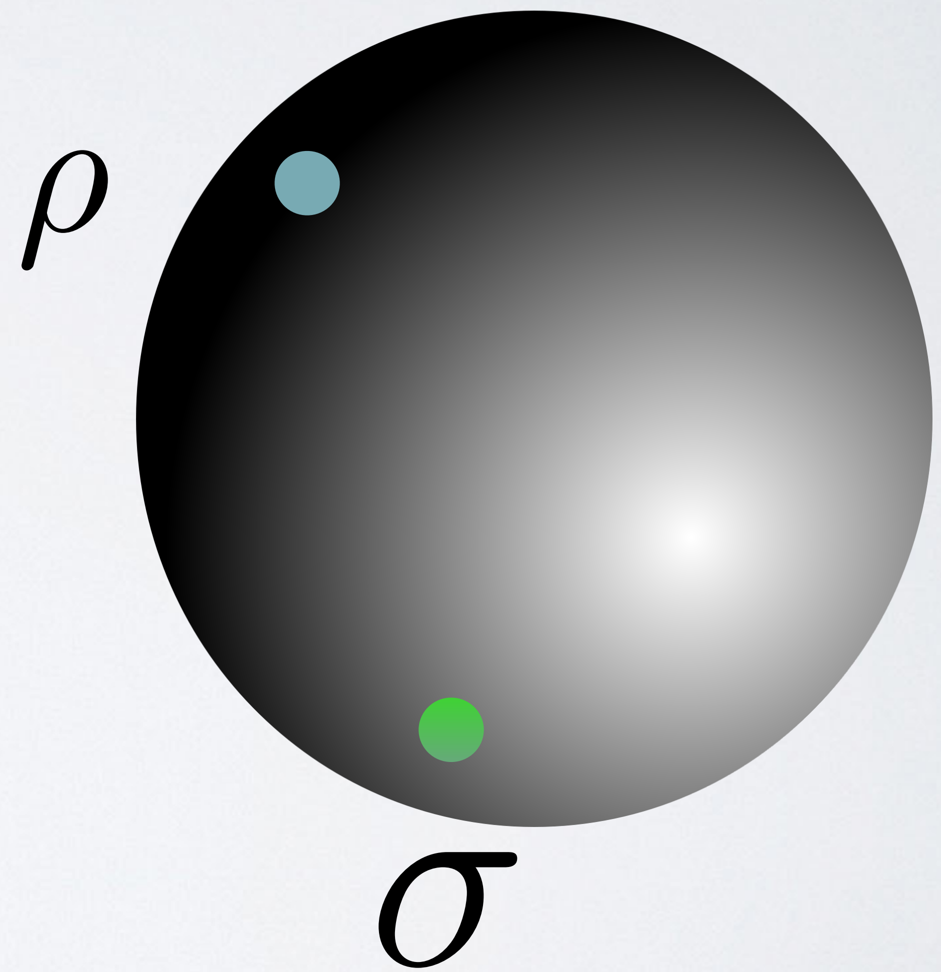
Which state is it?

Choose the higher likelihood!

$$\lambda(\rho, \sigma) = -2\log\left(\frac{\mathcal{L}(\rho)}{\mathcal{L}(\sigma)}\right)$$

$$\sigma \iff \lambda \geq 0$$

Neyman-Pearson lemma tells us
this is the most powerful test.



Quantum information makes connections to statistical inference in many ways.

State tomography

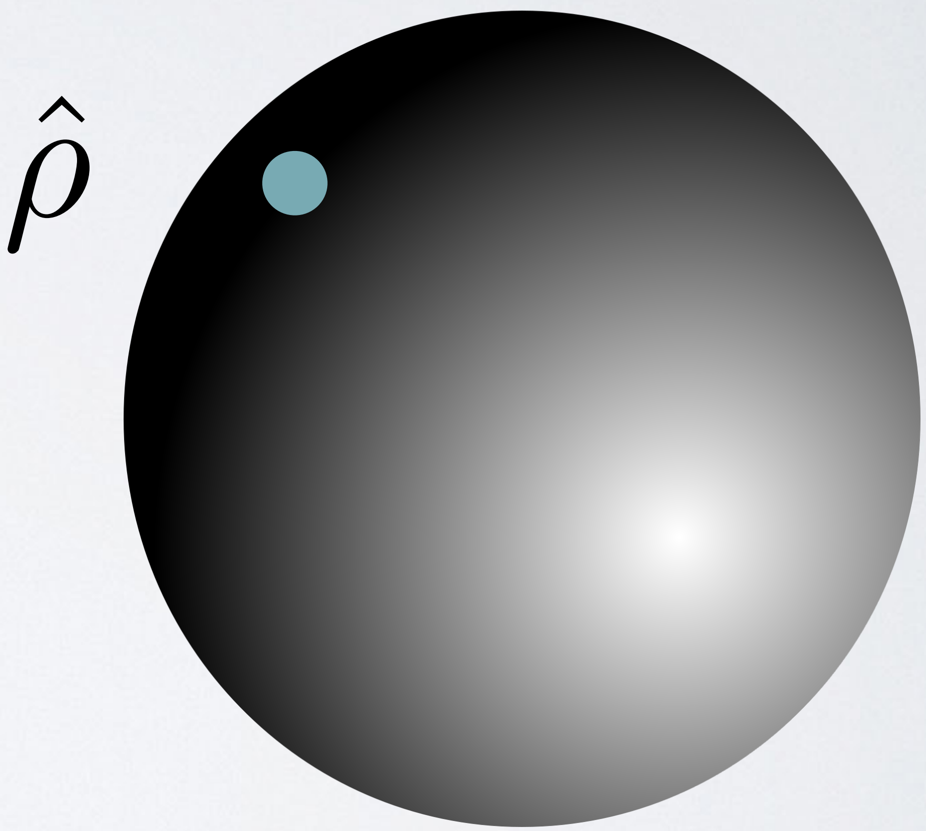
is an instance of

model fitting

Which parameters are best?

Maximum likelihood estimation

$$\hat{\rho} = \underset{\rho}{\operatorname{argmax}}\ \mathcal{L}(\rho)$$

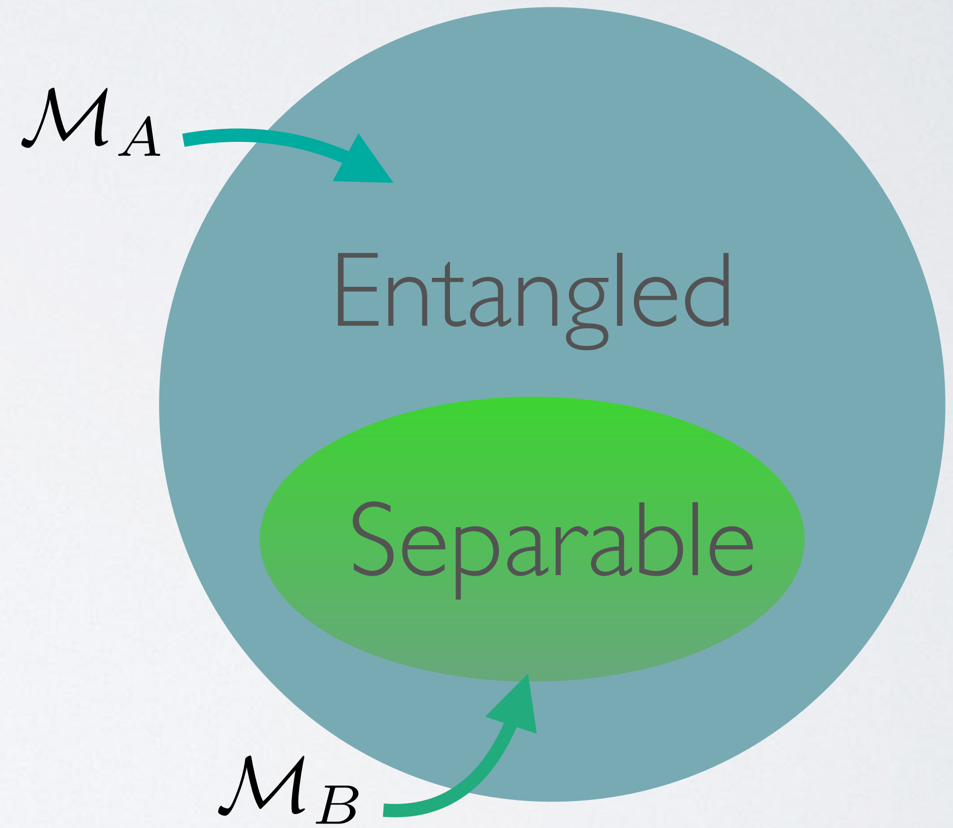


Quantum information makes connections to statistical inference in many ways.

Entanglement verification

is an instance of

composite hypothesis testing



Which region is it?

Choose the higher likelihood!

$$\lambda(\mathcal{M}_A, \mathcal{M}_B) = -2\log\left(\frac{\mathcal{L}(\mathcal{M}_A)}{\mathcal{L}(\mathcal{M}_B)}\right)$$

$$\mathcal{M}_B \iff \lambda \geq 0$$

Quantum information makes connections to statistical inference in many ways.

Z-diagonal state vs not

is an instance of

(nested) model selection

Is the true state on the line

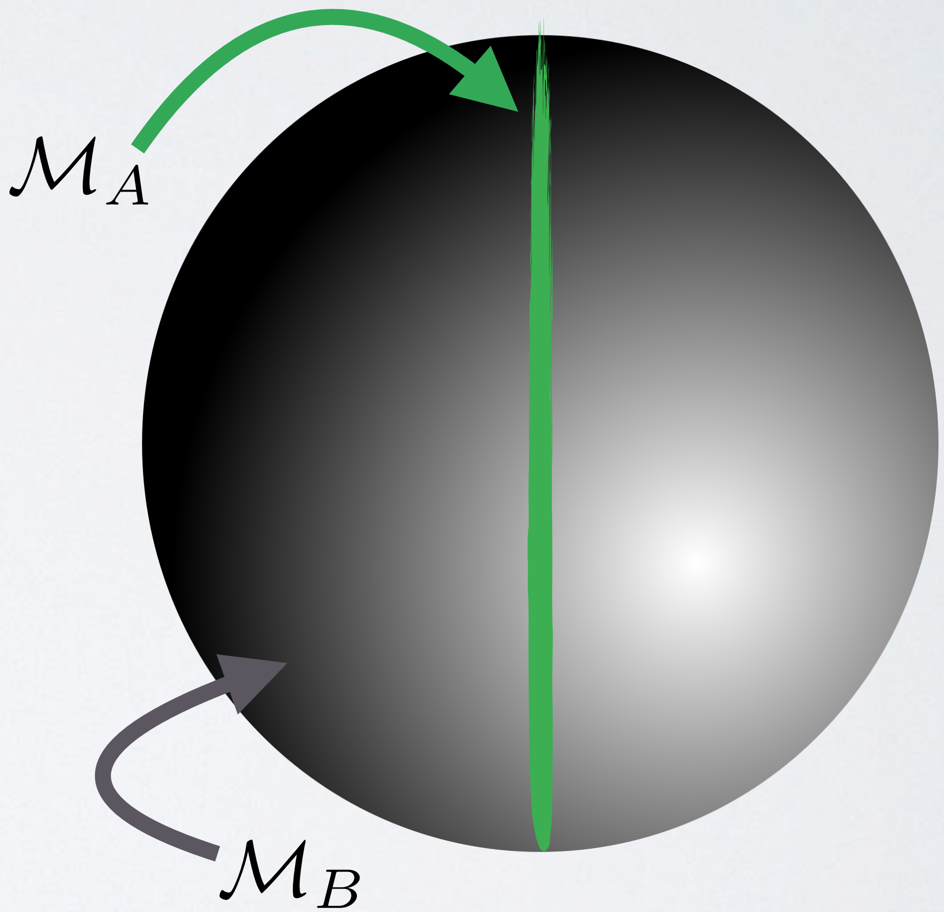
or not?

~~Choose the higher likelihood!~~

$$\lambda(\mathcal{M}_A, \mathcal{M}_B) = -2\log\left(\frac{\mathcal{L}(\mathcal{M}_A)}{\mathcal{L}(\mathcal{M}_B)}\right)$$

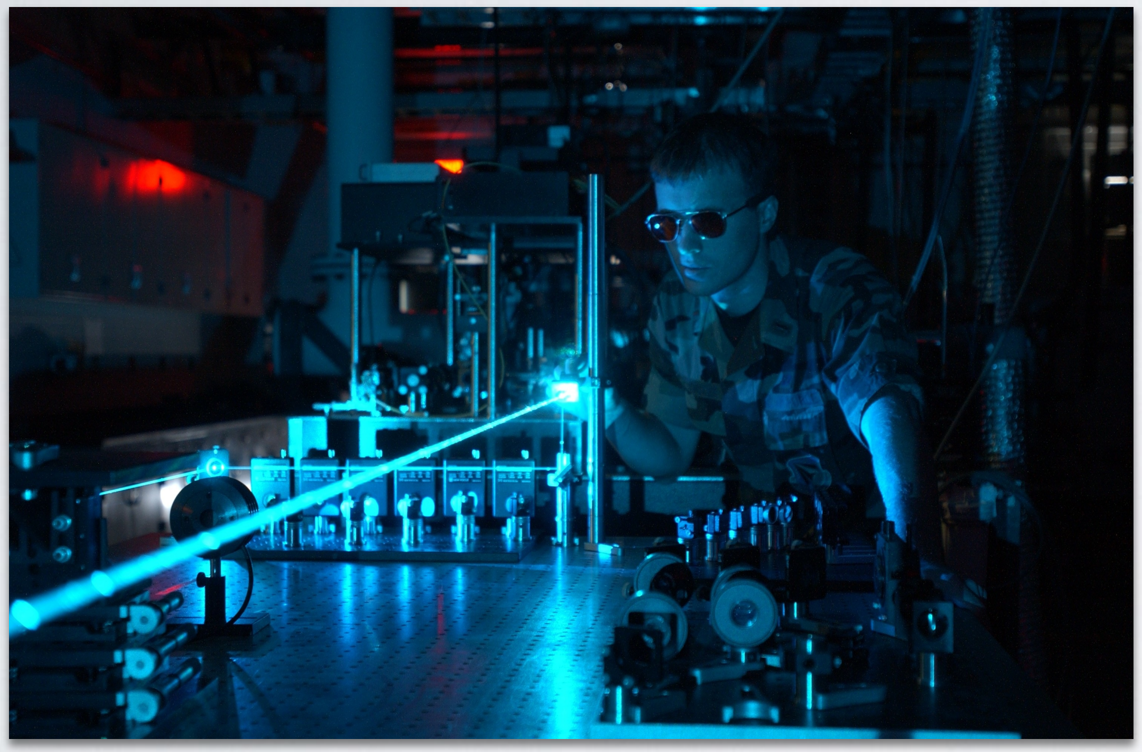
$$\lambda \geq 0 \implies \mathcal{M}_B\ ???$$

LLRS never negative!

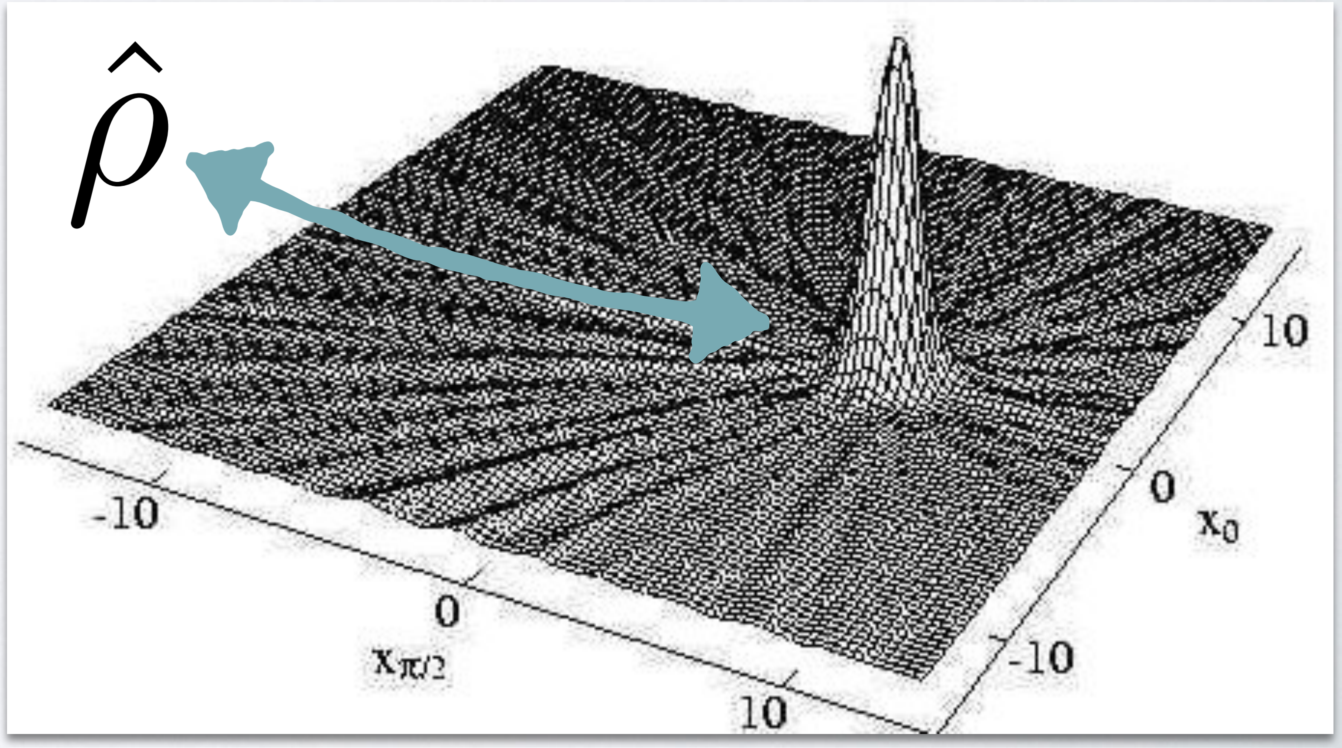


How will we **investigate** model selection in state tomography?

I am studying a paradigmatic problem: tomography of continuous-variable systems.



Optical modes of light….

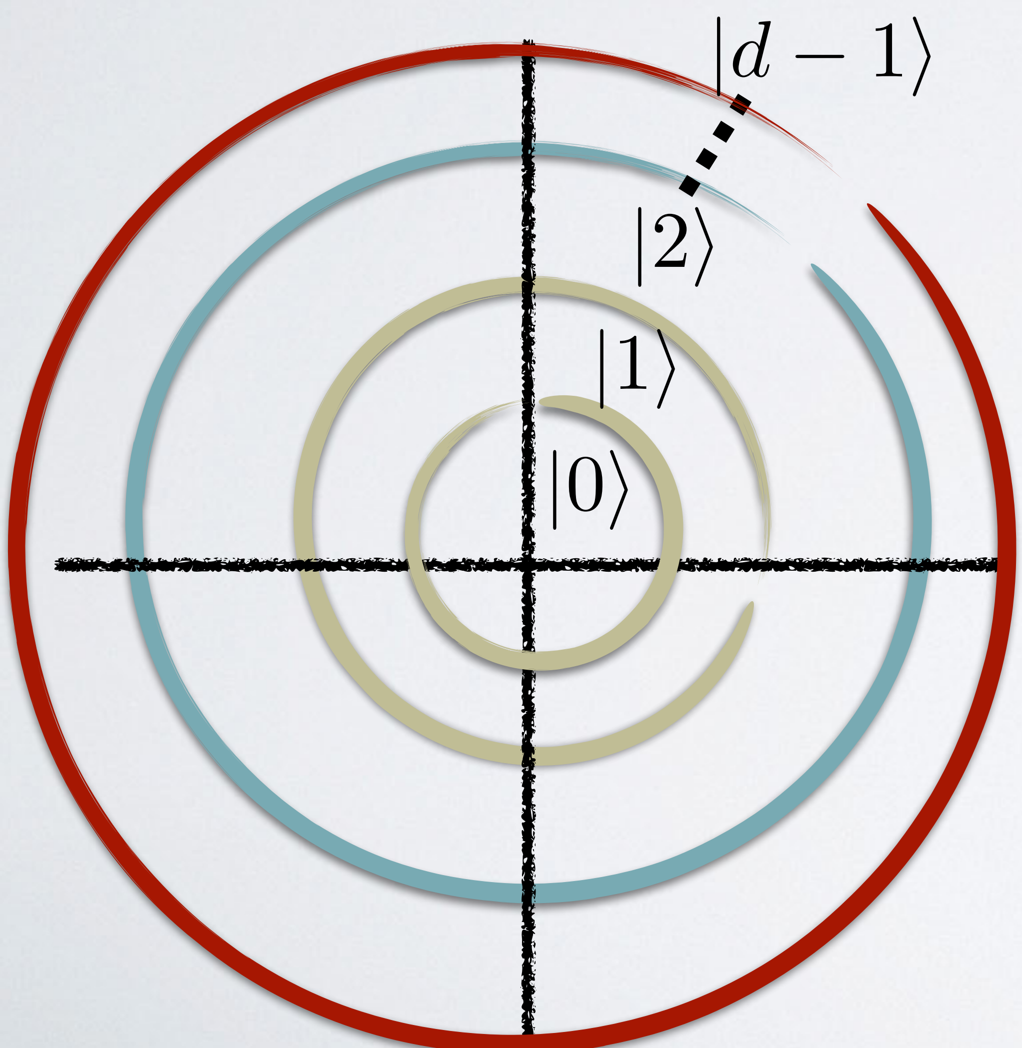


…as *Wigner functions or density matrices.*

The models I consider are *subspaces* of an infinite-dimensional Hilbert space.

$$\mathcal{H}_d = \mathrm{Span}\left(|0\rangle, |1\rangle, \cdots |d-1\rangle\right)$$

$$\mathcal{M}_d = \{\rho \mid \rho \in \mathcal{B}(\mathcal{H}_d),\ Tr(\rho) = 1,\ \rho \geq 0\}$$



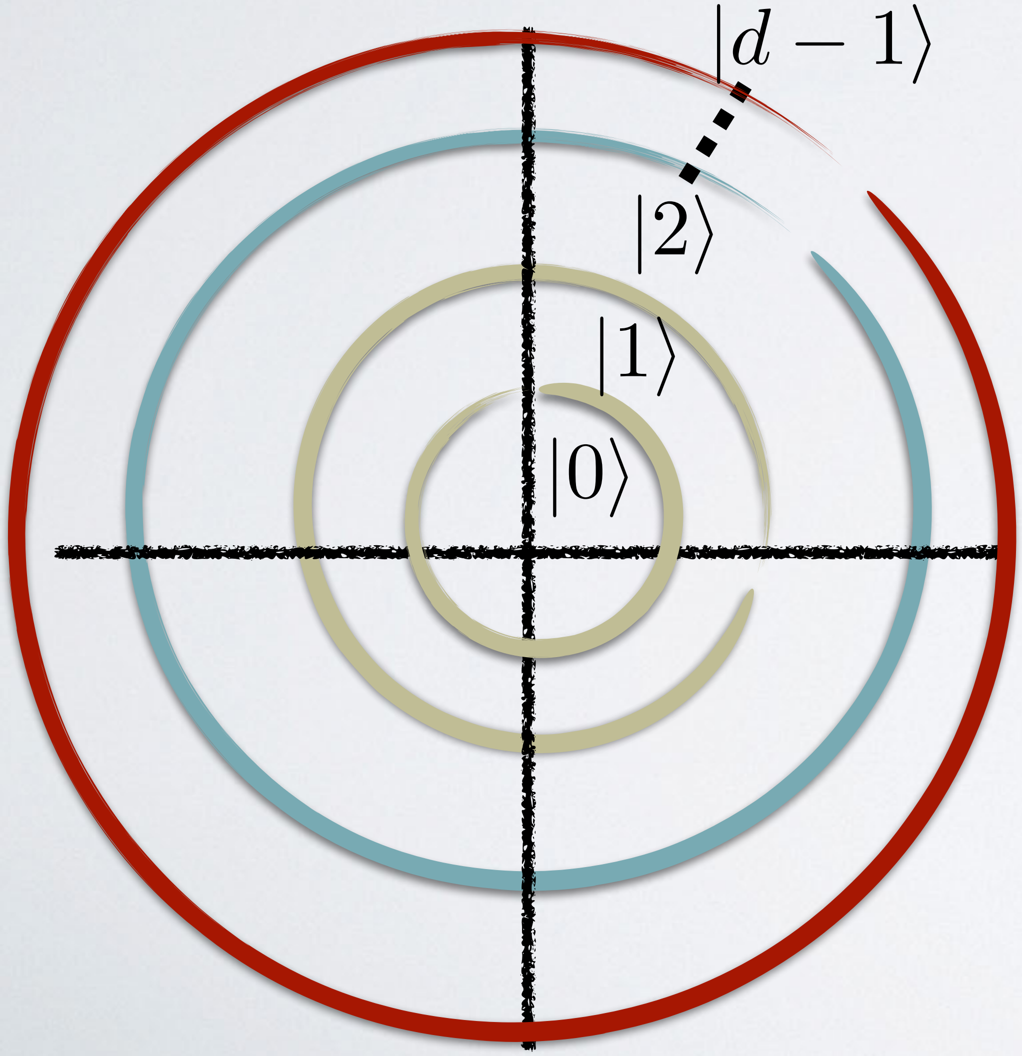
Models come from low-energy assumption (and lack of structure in Wigner function)

Other models are possible (e.g., by *rank*).

The models I consider are *nested* inside one another.

$$\mathcal{M}_d \subset \mathcal{M}_{d+1}$$

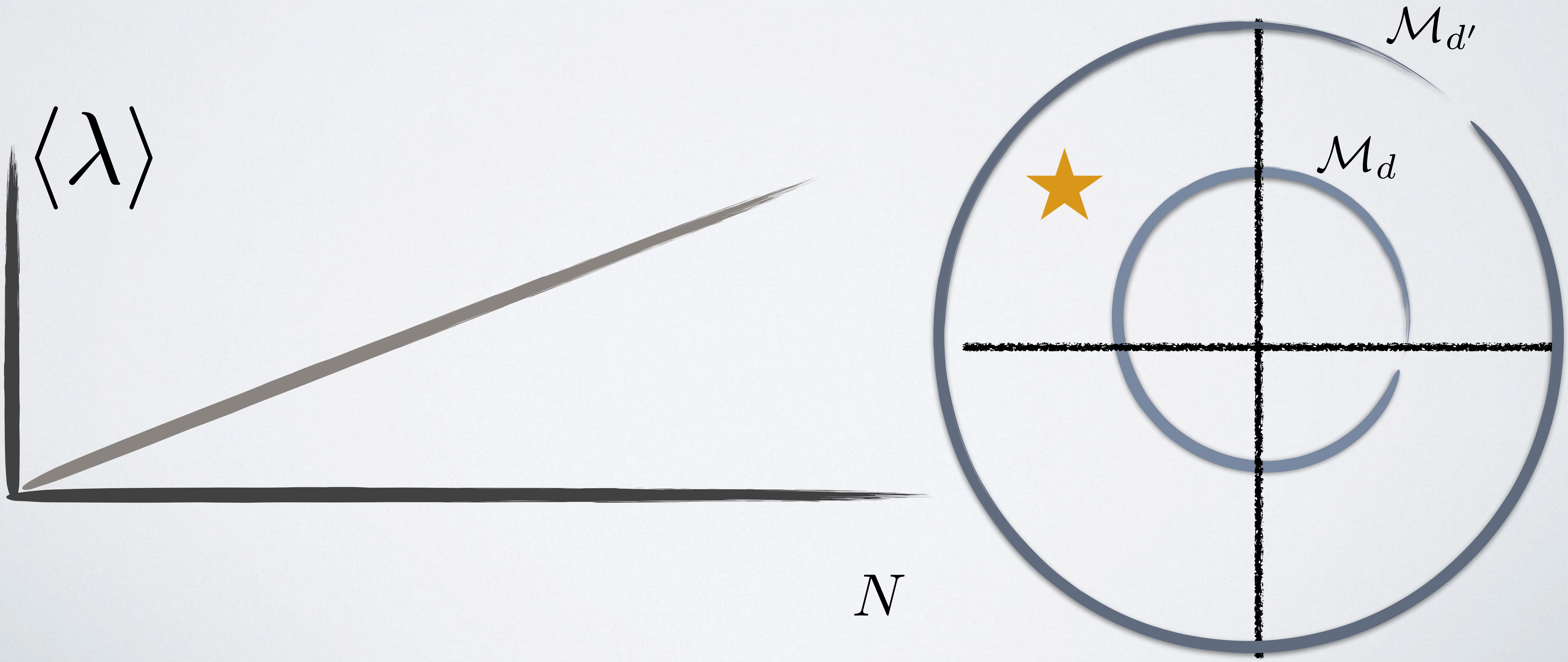
How can we use likelihoods to compare them?



We have to tackle nested model selection

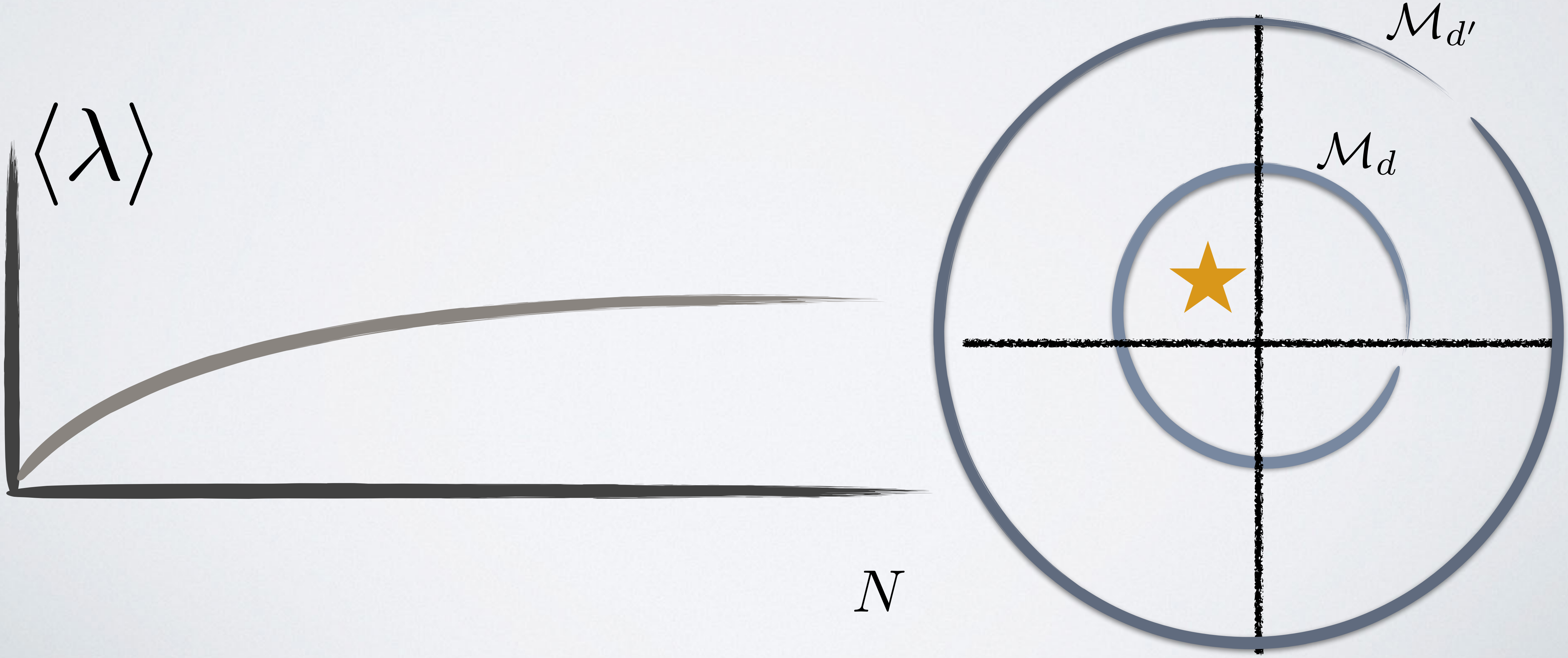
We re-think the use of the LLRS for nested model selection, based on its expected value.

$$\lambda(\mathcal{M}_d, \mathcal{M}_{d'}) = -2 \log \left(\frac{\mathcal{L}(\mathcal{M}_d)}{\mathcal{L}(\mathcal{M}_{d'})} \right)$$



We re-think the use of the LLRS for nested model selection, based on its expected value.

$$\lambda(\mathcal{M}_d, \mathcal{M}_{d'}) = -2\log\left(\frac{\mathcal{L}(\mathcal{M}_d)}{\mathcal{L}(\mathcal{M}_{d'})}\right)$$



Asymptotic convergence of the LLRS is a consequence of the *Wilks Theorem*.

THE LARGE-SAMPLE DISTRIBUTION OF THE LIKELIHOOD RATIO FOR TESTING COMPOSITE HYPOTHESES[1]

BY S. S. WILKS

1938: Wilks gives distribution of LLRS.

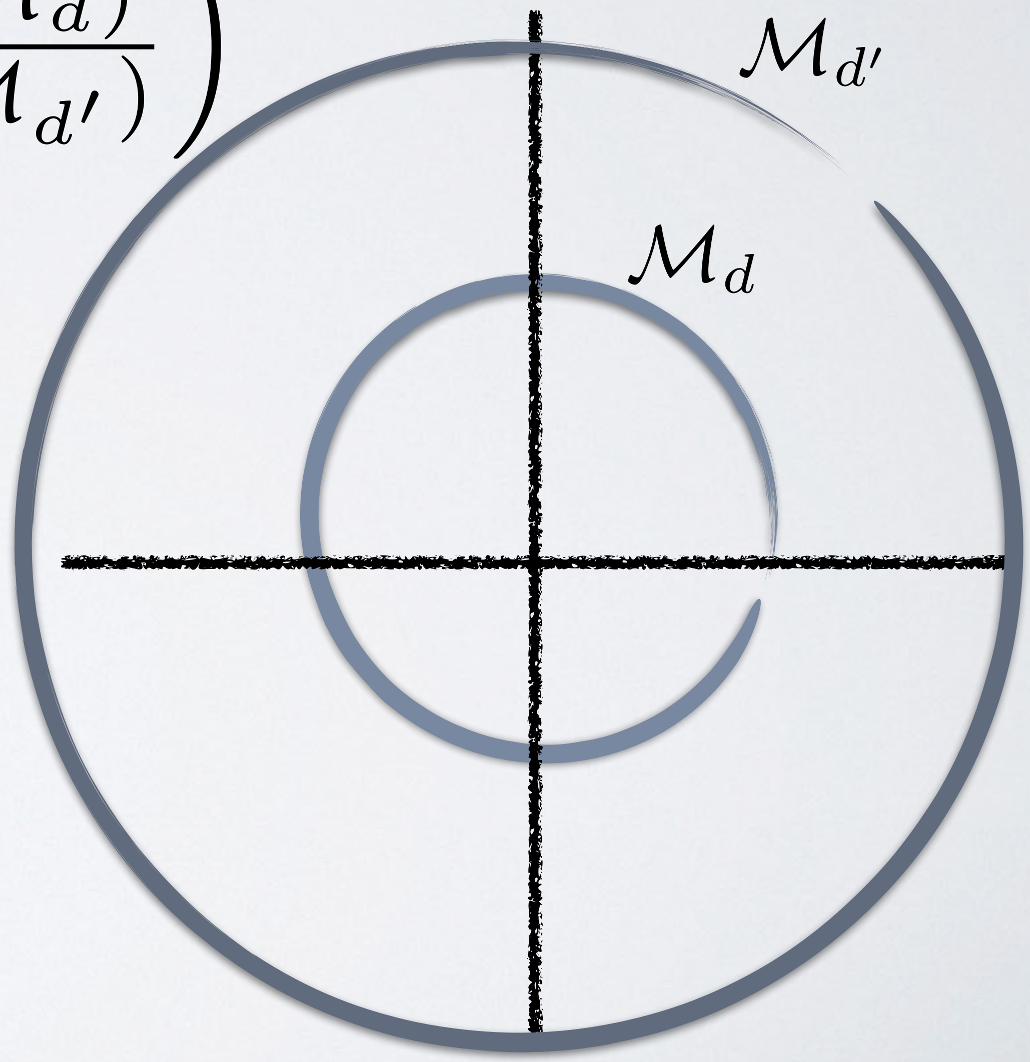
$$\lambda(\mathcal{M}_d, \mathcal{M}_{d'}) \sim \chi^2_{p_{d'} - p_d}$$

The Wilks Theorem allows us to do nested model selection.

We compare the LLRS to its *expected value*

$$\lambda(\mathcal{M}_d, \mathcal{M}_{d'}) = -2\log\left(\frac{\mathcal{L}(\mathcal{M}_d)}{\mathcal{L}(\mathcal{M}_{d'})}\right)$$

$$\mathcal{M}_{d'} \iff \lambda \geq \langle\lambda\rangle$$



Another model selection technique relies on this result.

Information criteria **explicitly trade off** between fitting data well and having high accuracy

Use of *Akaike's* ***AIC*** is **common**

Relies on Wilks to compute bias of a particular estimator of KL divergence

My work feeds into creating a quantum information criterion

We now have **potential tools** for **nested model selection** in tomography. How do they **perform**?

I performed a Monte Carlo study of the LLRS and its behavior.

Studied:

- 17 true states (supported on low-energy subspace)

I performed a Monte Carlo study of the LLRS and its behavior.

Studied:

- 17 true states
- 100 random datasets for each state (*coherent state POVM*)

$$\mathrm{Data} = \{\alpha_j \mid \alpha_j \in \mathbb{C},\ Pr(\alpha_j) = \langle\alpha_j|\rho_{true}|\alpha_j\rangle\}$$

I performed a Monte Carlo study of the LLRS and its behavior.

Studied:

- 17 true states
- 100 random datasets for each state
- 10K to 100K samples for each dataset

I performed a Monte Carlo study of the LLRS and its behavior.

Studied:

- 17 true states
- 100 random datasets for each state
- 10K to 100K samples for each dataset
- MLE over {2…10}-dimensional Hilbert spaces

$$\{\mathcal{M}_2, \mathcal{M}_3, \cdots, \mathcal{M}_{10}\}$$

Lots of supercomputer time!

The results were **puzzling**.

I checked *four predictions* of the Wilks Theorem on the behavior of the LLRS. Only *one matched*.

THE LARGE-SAMPLE DISTRIBUTION OF THE LIKELIHOOD RATIO FOR TESTING COMPOSITE HYPOTHESES[1]

BY S. S. WILKS

Predictions:

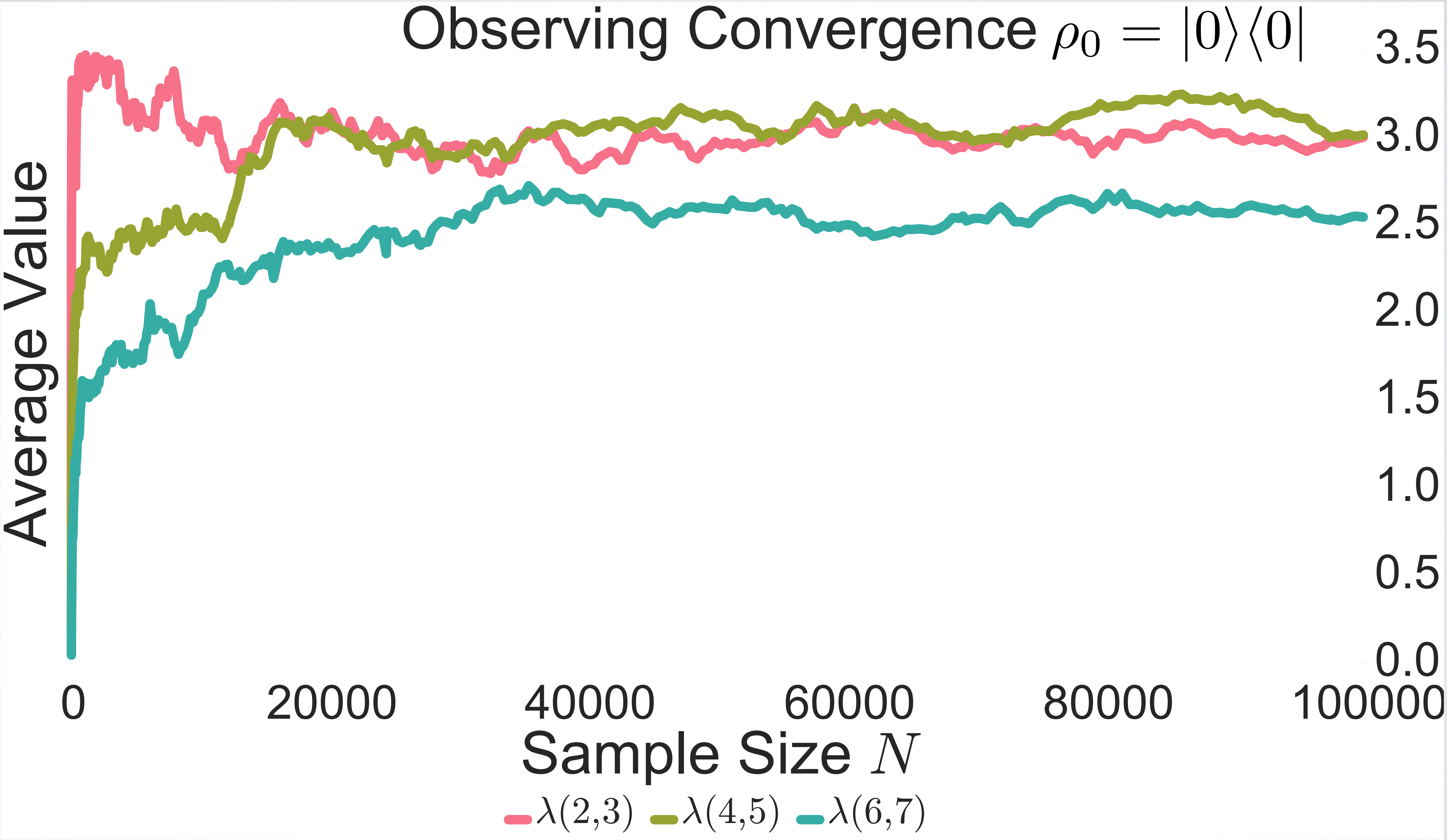
Asymptotic convergence

A particular expected value

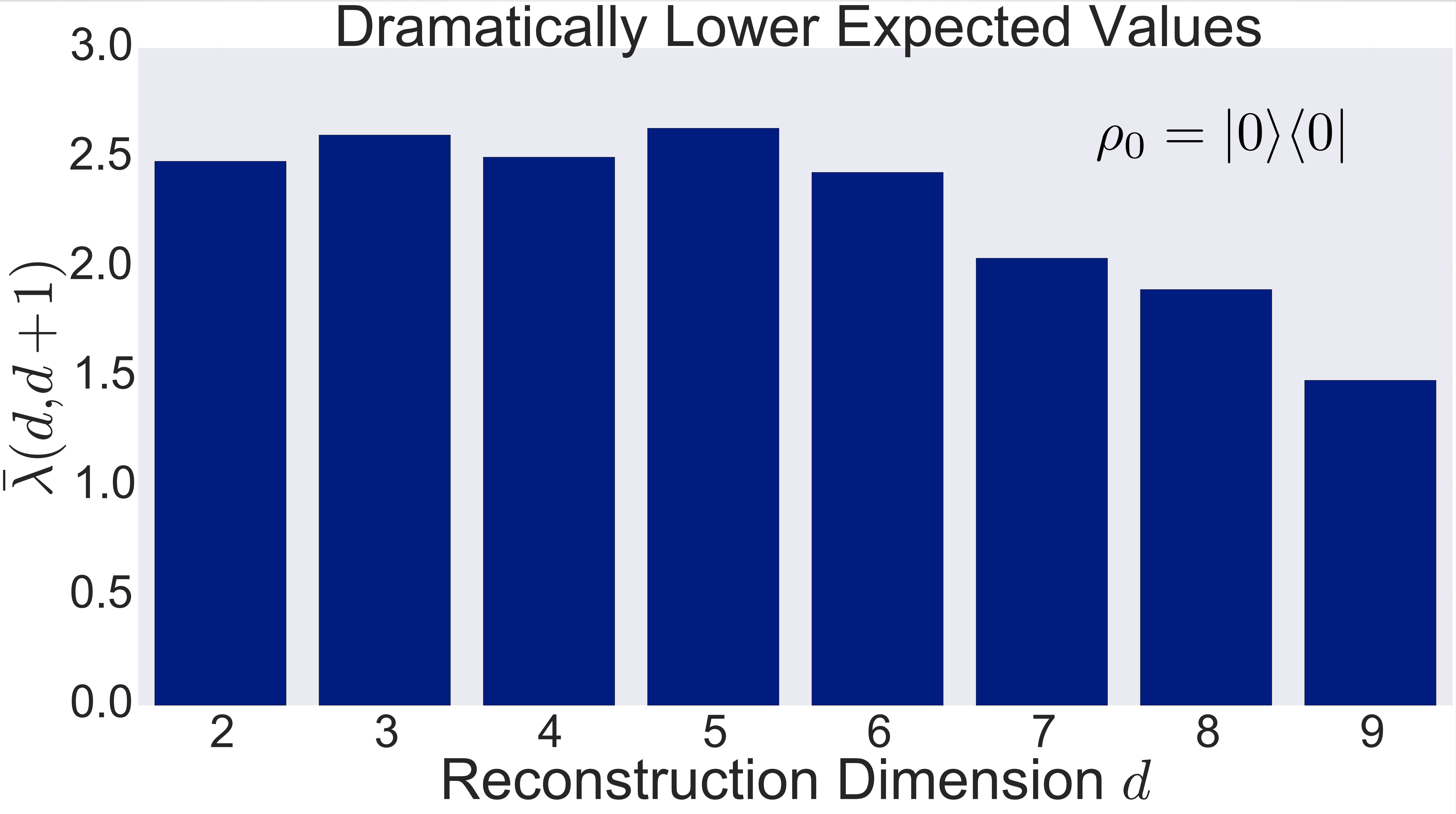
Distribution independent of truth

Distribution depends on reconstruction dimension

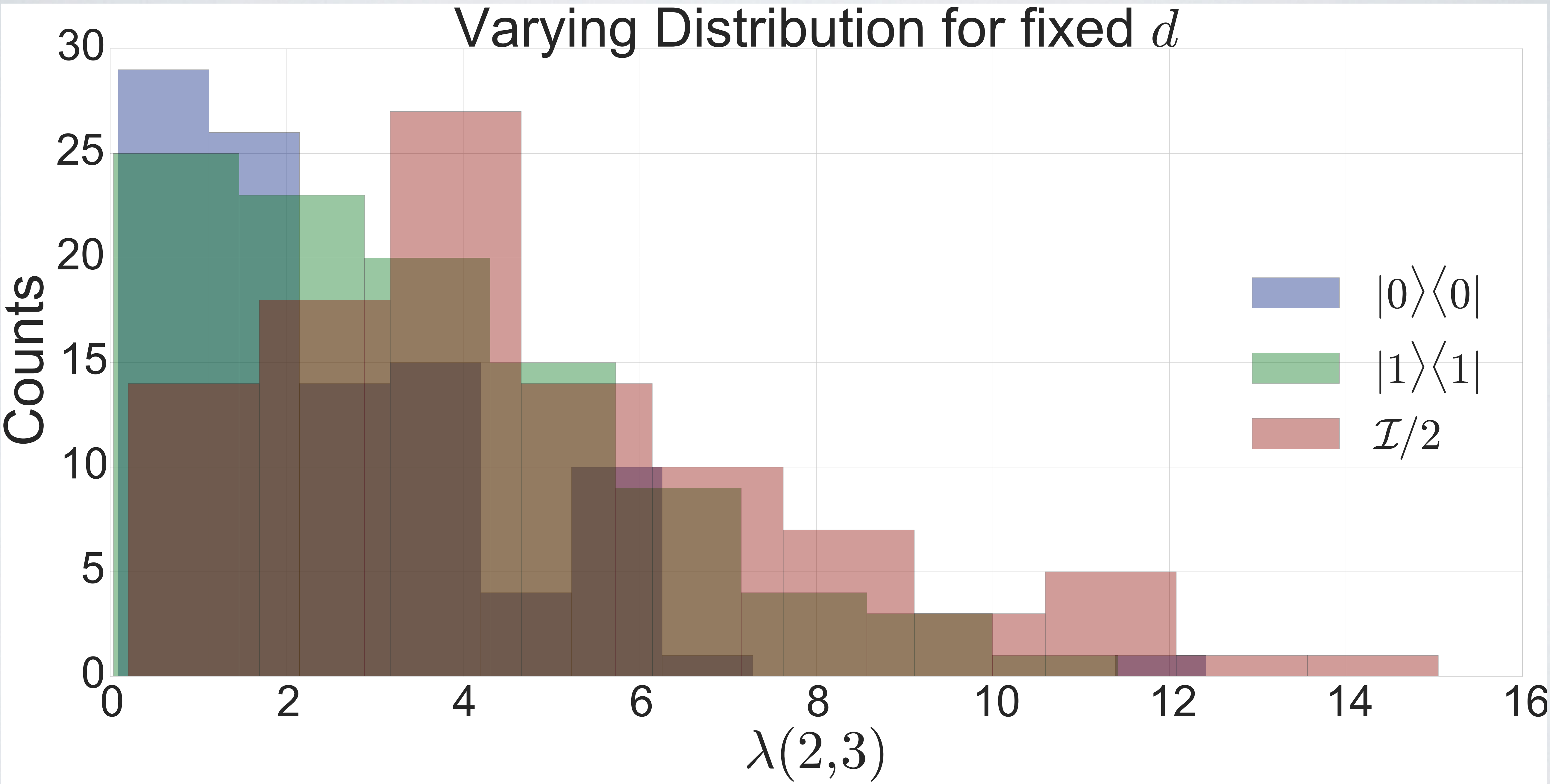
When the truth is in the smaller model, we observe asymptotic convergence.



Monte Carlo averages and expectation values do not agree at all.

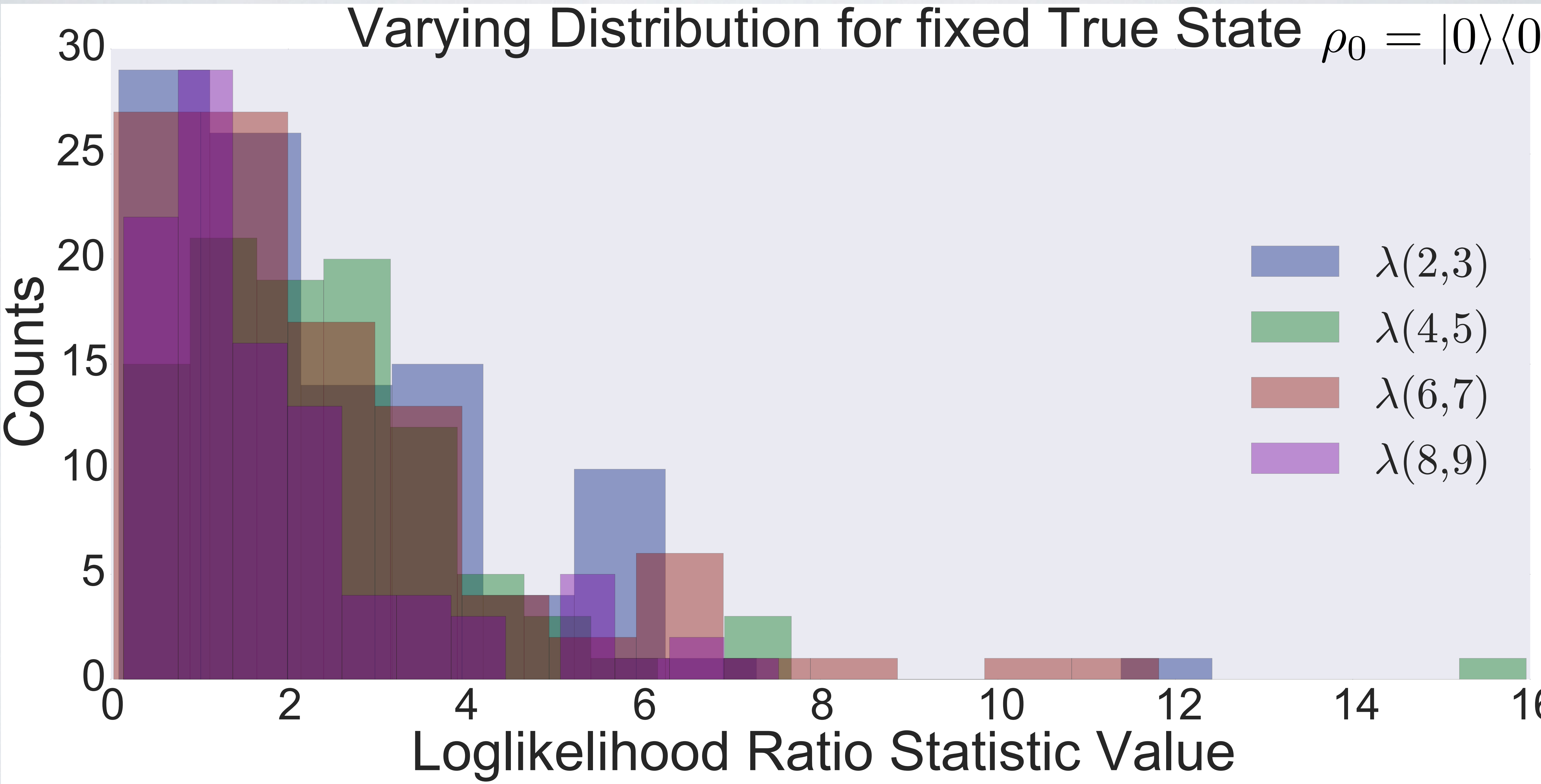


Wilks theorem predictions for distribution of LLRS do not agree with simulation.



Wilks: Distribution *independent of true state*

Wilks theorem predictions for distribution of LLRS do not agree with simulation.



Wilks: Distribution *depends on reconstruction dimension*

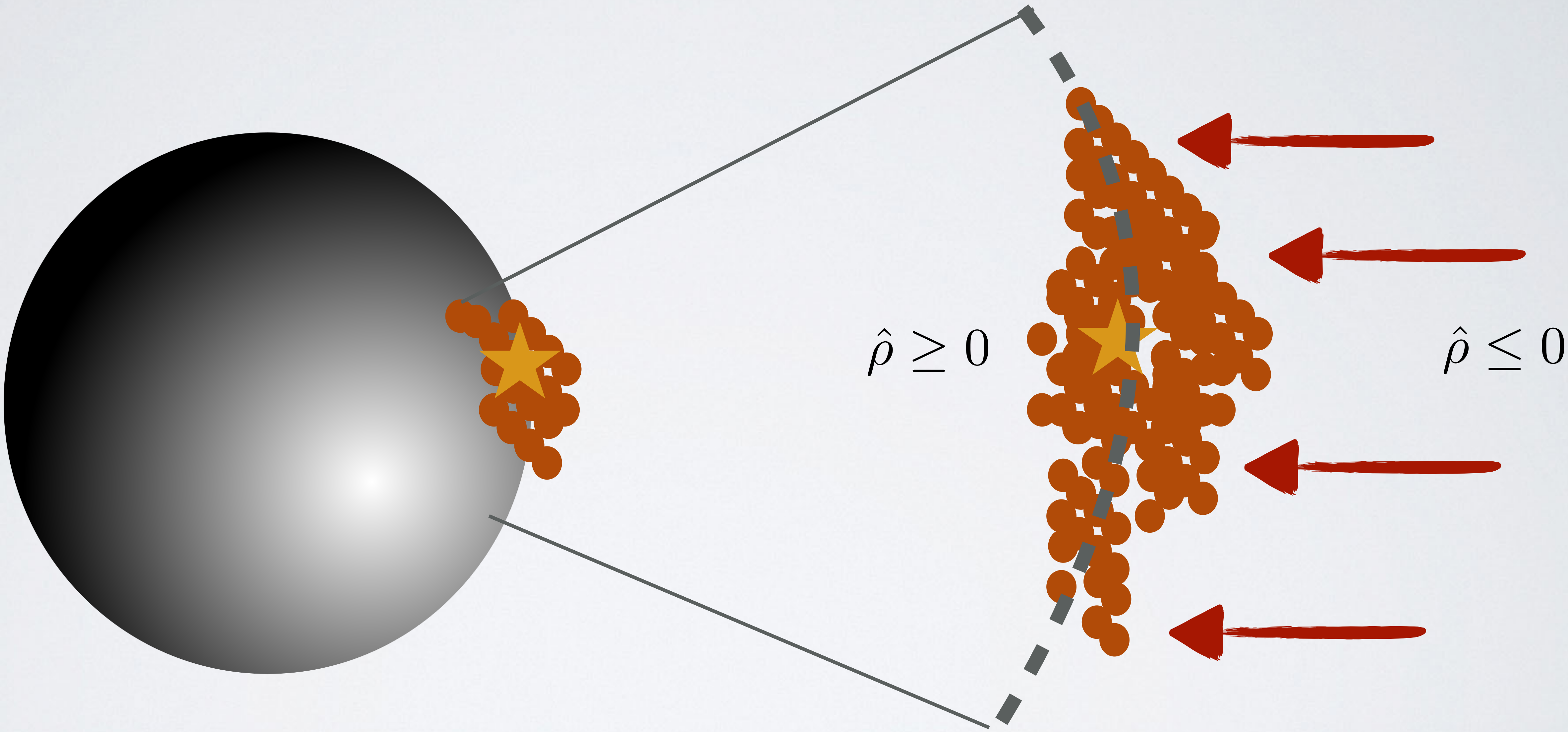
Theorems are not "wrong", only "not applicable".

Why does the Wilks Theorem **not apply**?

State tomography is **on the edge**.

Let’s see why.

The first edge is the positivity constraint. This shows up a lot in quantum information.

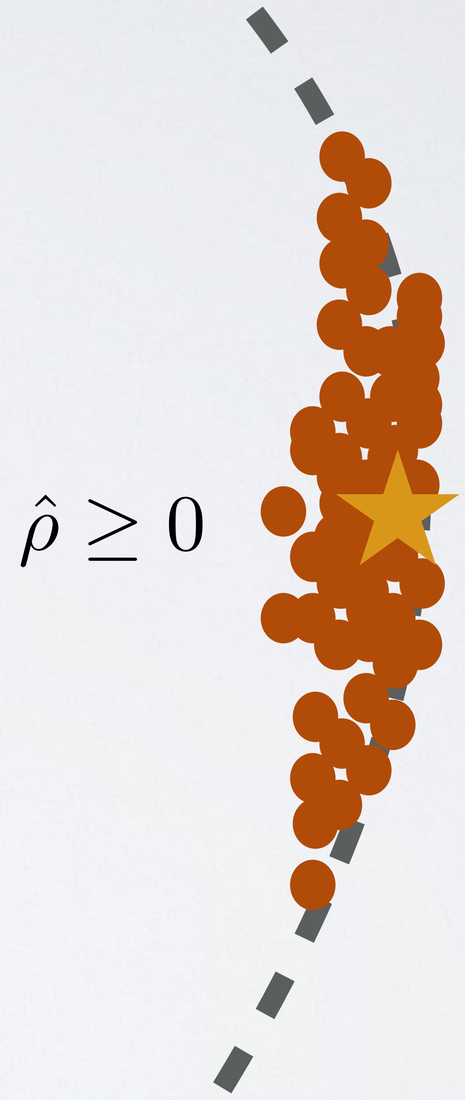


The first edge is the positivity constraint. This shows up a lot in quantum information.

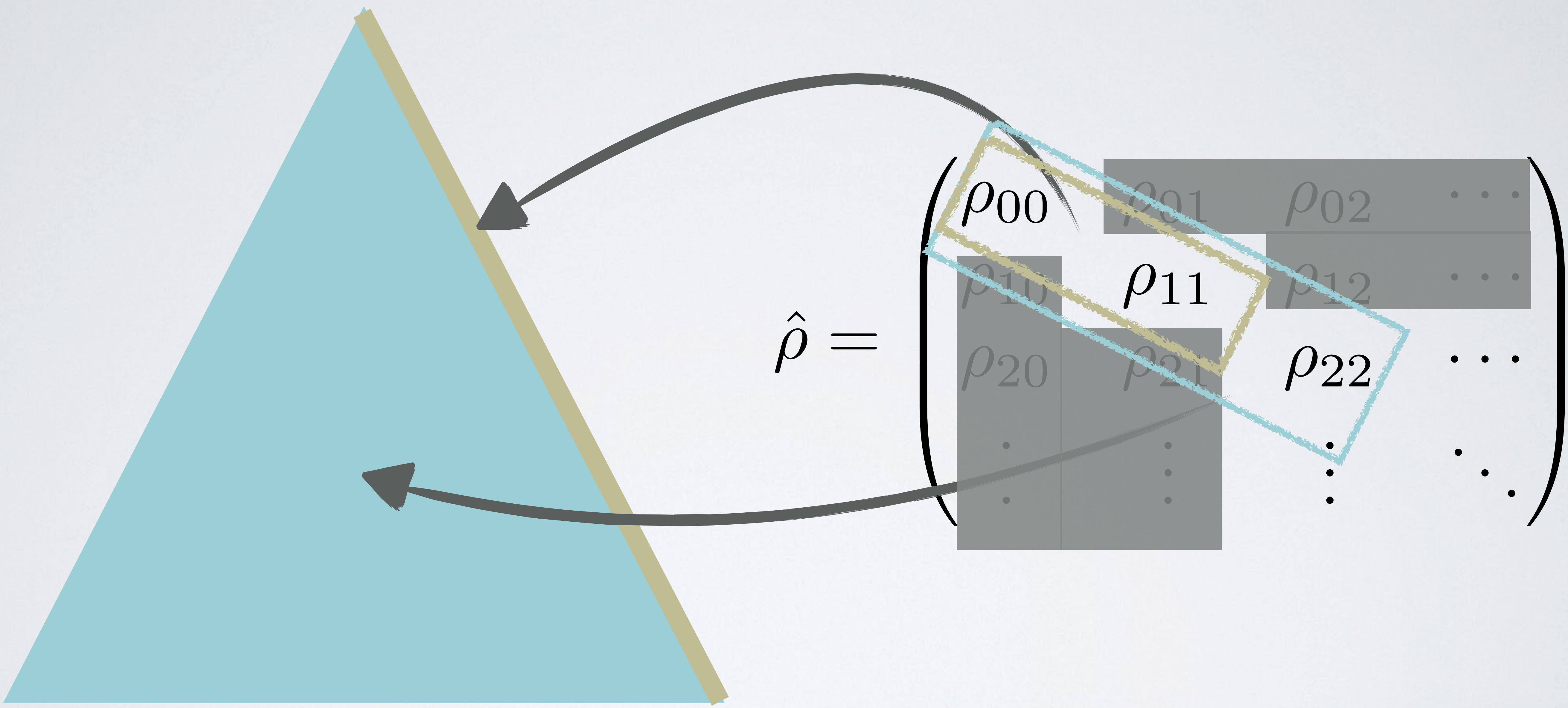
Positivity "piles up" estimates on the boundary

Fluctuations normal to boundary are *diminished*.

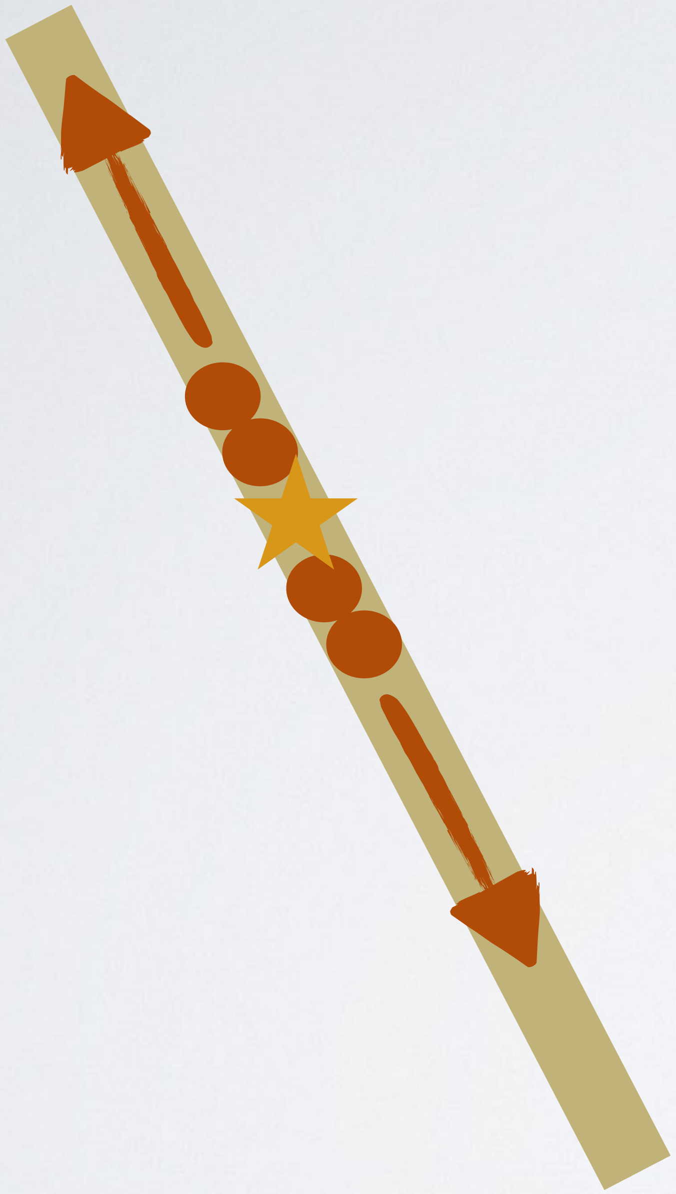
Results in *bias* of the estimator

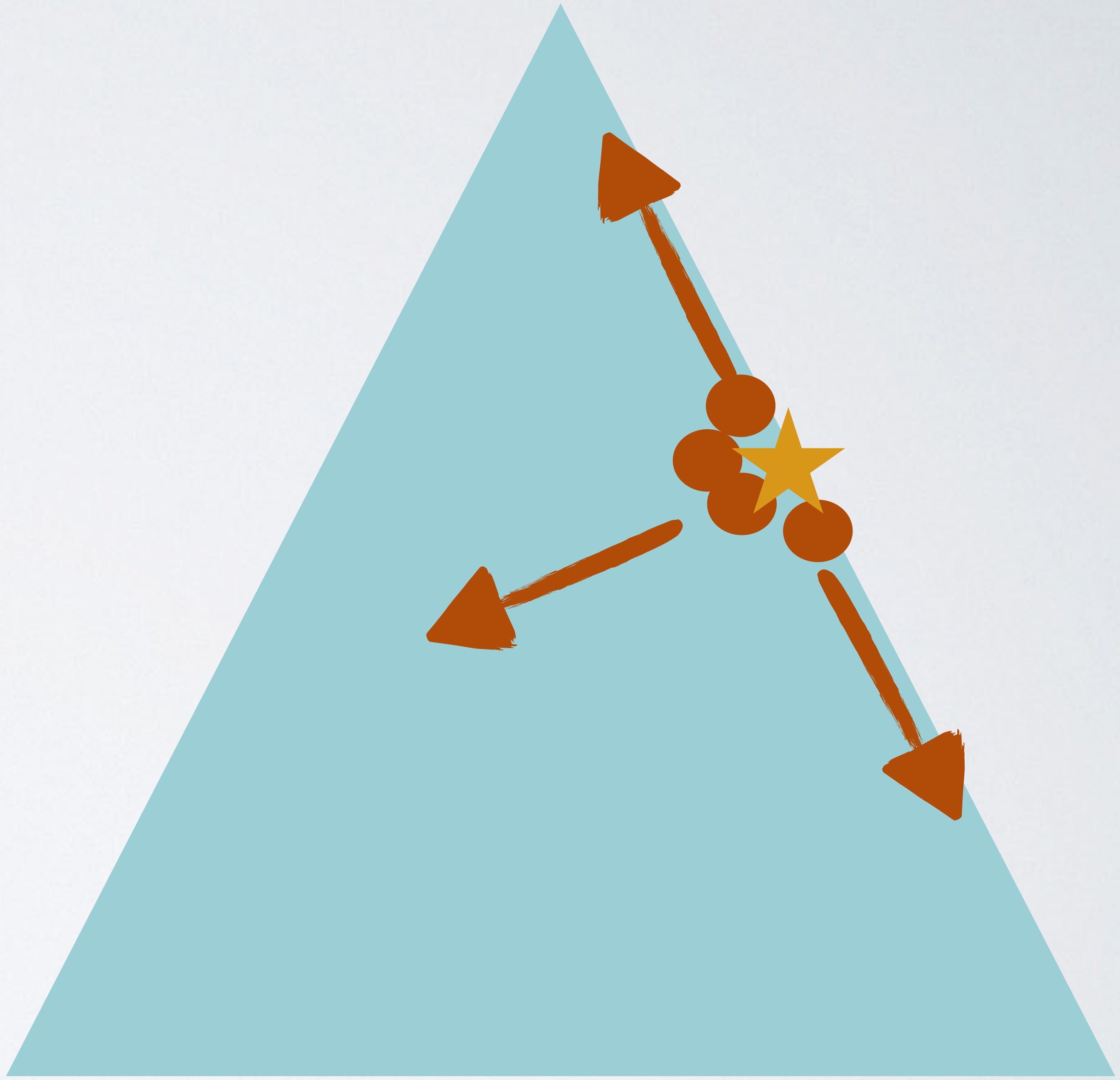


The second edge is that the Hilbert spaces I use nest on the boundary of one another.

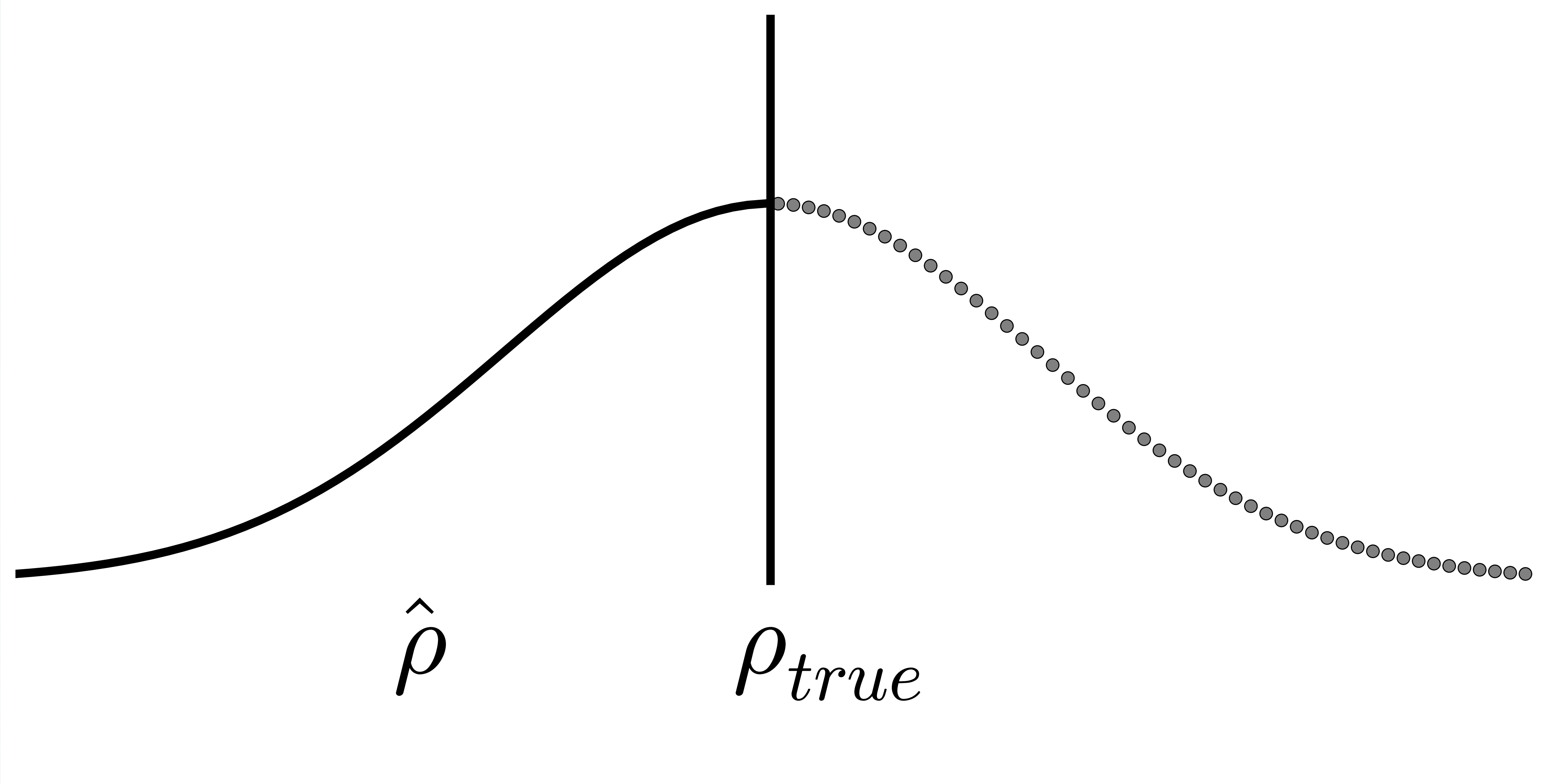


When the true state is mixed, you avoid the first edge, but still run right into the second.



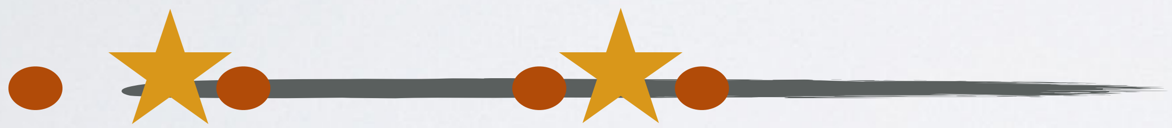


The Wilks Theorem cannot be applied on boundaries - they introduce constraints.



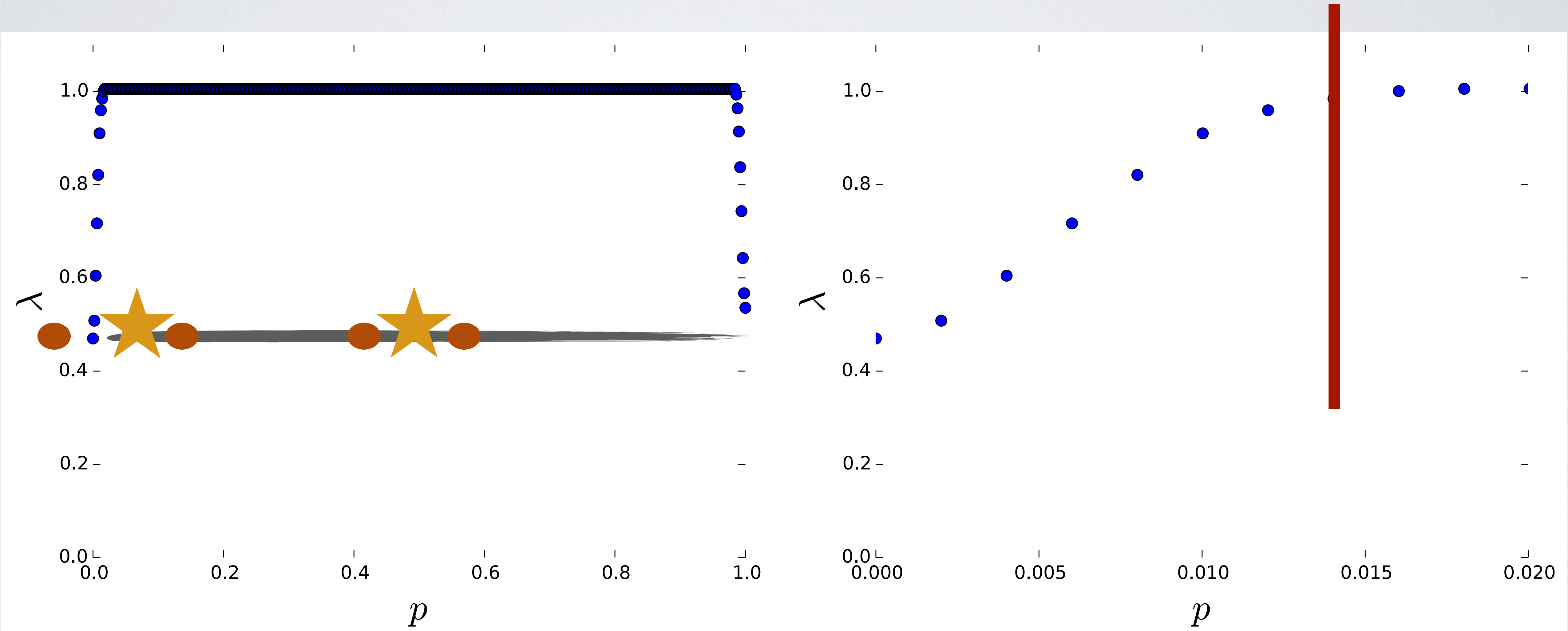
Boundaries change distribution of MLEs, causing problems.

A numerical study of a related problem indicates closeness to the boundary affects the LLRS.



1-simplex (i.e., a coin)

A numerical study of a related problem indicates closeness to the boundary affects the LLRS.



If truth within 1 standard deviation of boundary, constraints become important.

State tomography is **on the edge**.

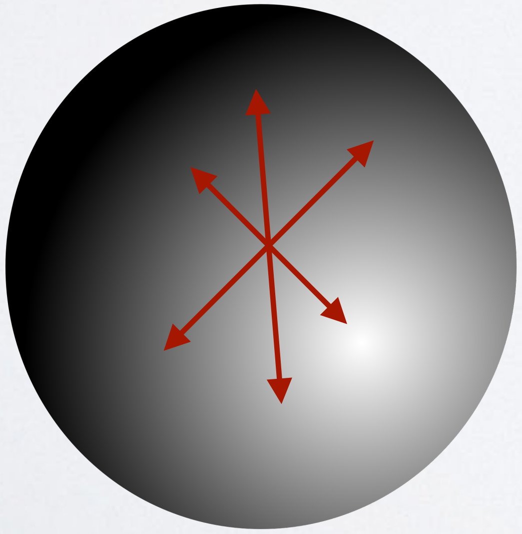
So must our model selection be.

Proving a "qWilks theorem" would be hard, in general.

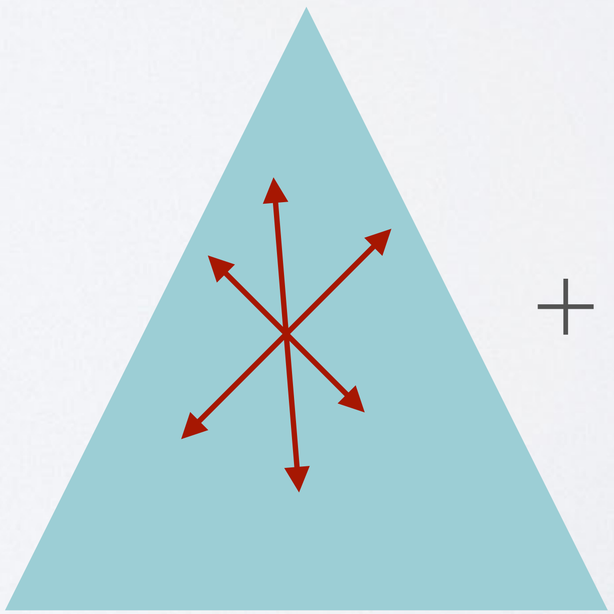
Distribution of lambda depends on true state

Distribution depends on Hilbert space dimension

Quantum state space hard to reason about…



3



+ 6

Can we find a replacement for the Wilks theorem which respects boundaries?

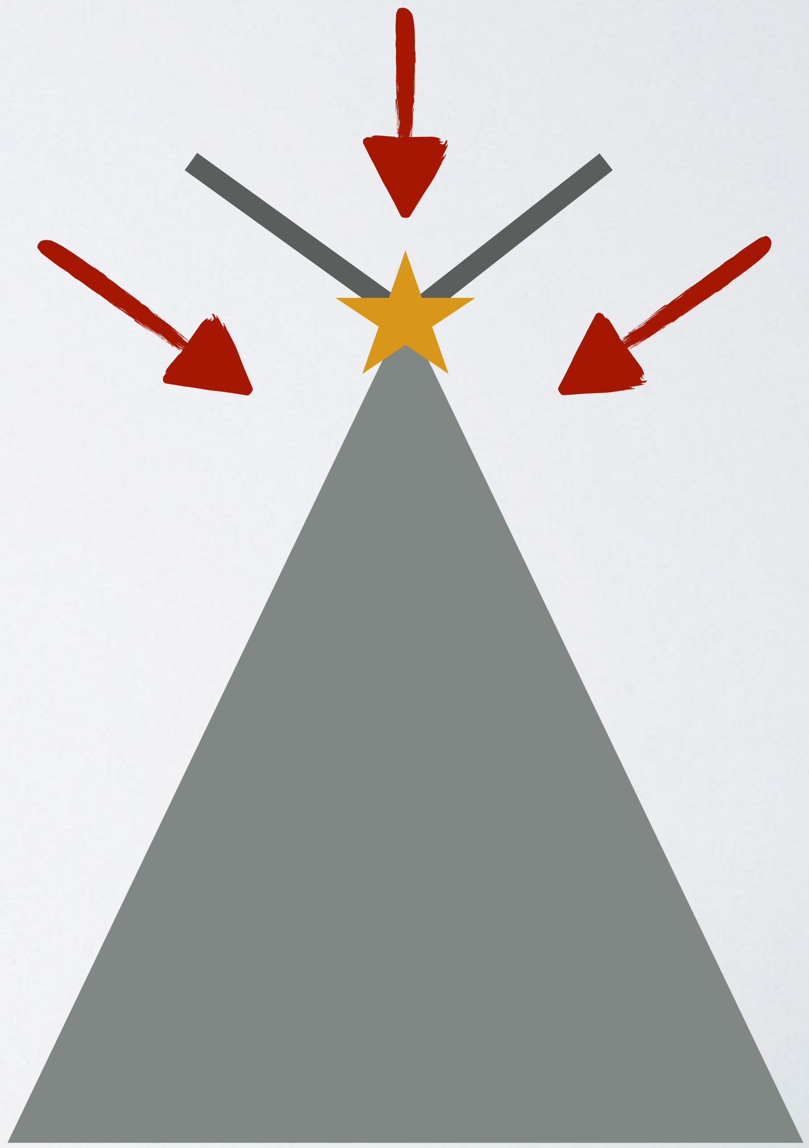
Quantum states = classical simplexes + unitary DOF

LLRS depends on *rank of true state*

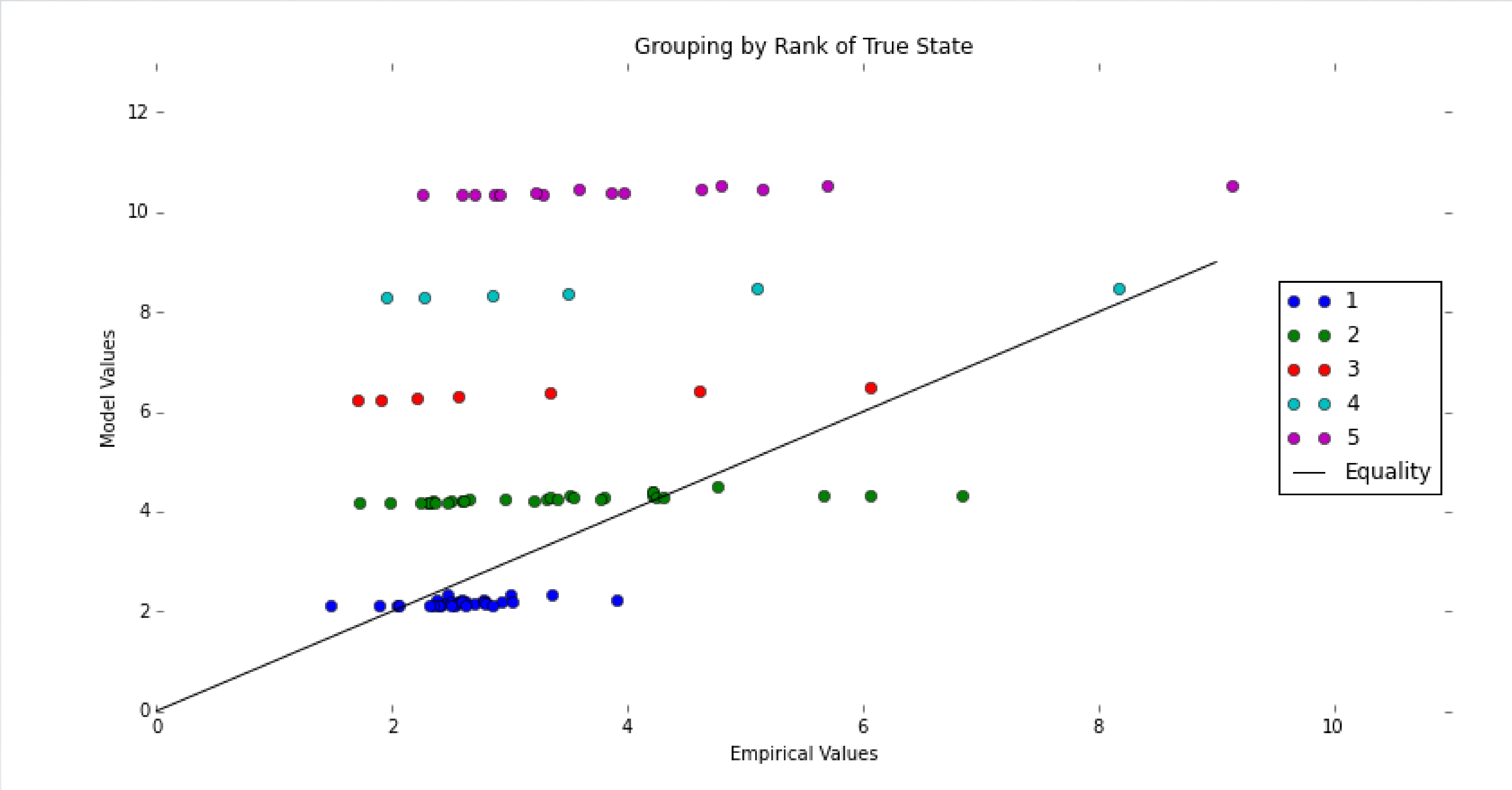
Does require Monte Carlo for simulating effect of simplex boundaries

$$\langle \lambda \rangle = 2\ \mathrm{rank}(\rho_{true}) + f(d+1, r) - f(d, r)$$

	1	2	3	4	5	6
2	0.496443	0.999158	NaN	NaN	NaN	NaN
3	0.822011	1.488624	1.982415	NaN	NaN	NaN
4	1.070604	1.875747	2.478096	2.976010	NaN	NaN
5	1.263654	2.189043	2.883159	3.465806	3.959832	NaN
6	1.442006	2.480977	3.266152	3.926972	4.485939	4.985161
7	1.599560	2.728116	3.589044	4.313688	4.933024	5.488181
8	1.731031	2.937367	3.865390	4.648795	5.319806	5.921281
9	1.846602	3.127363	4.116697	4.955233	5.675212	6.321233
10	1.962161	3.317975	4.365928	5.258693	6.024682	6.715674
11	2.076795	3.502036	4.610776	5.554809	6.365295	7.098042
12	2.171777	3.666289	4.831894	5.828708	6.685554	7.458987



How well does this replacement work?



Not as accurate as we expected…what is going on?

When in doubt, do a Taylor series expansion of something!

Helpful trick: $$\lambda(\mathcal{M}_d, \mathcal{M}_{d'}) = \lambda(\rho_{true}, \mathcal{M}_{d'}) - \lambda(\rho_{true}, \mathcal{M}_d)$$

How does LLRS change *when truth and estimates are close?*

Expand LLRS *as function of true state* to second order

The Taylor series allows us to calculate an approximate expected value.

$$\lambda(\rho_{true}, \mathcal{M}_d) = \lambda(\rho_{true}, \hat{\rho})$$

$$\approx 0 + \left.\frac{\partial \lambda}{\partial \rho}\right|_{\hat{\rho}} (\rho_{true} - \hat{\rho}) + \frac{1}{2} \left.\frac{\partial^2 \lambda}{\partial \rho^2}\right|_{\hat{\rho}} (\rho_{true} - \hat{\rho})^2$$

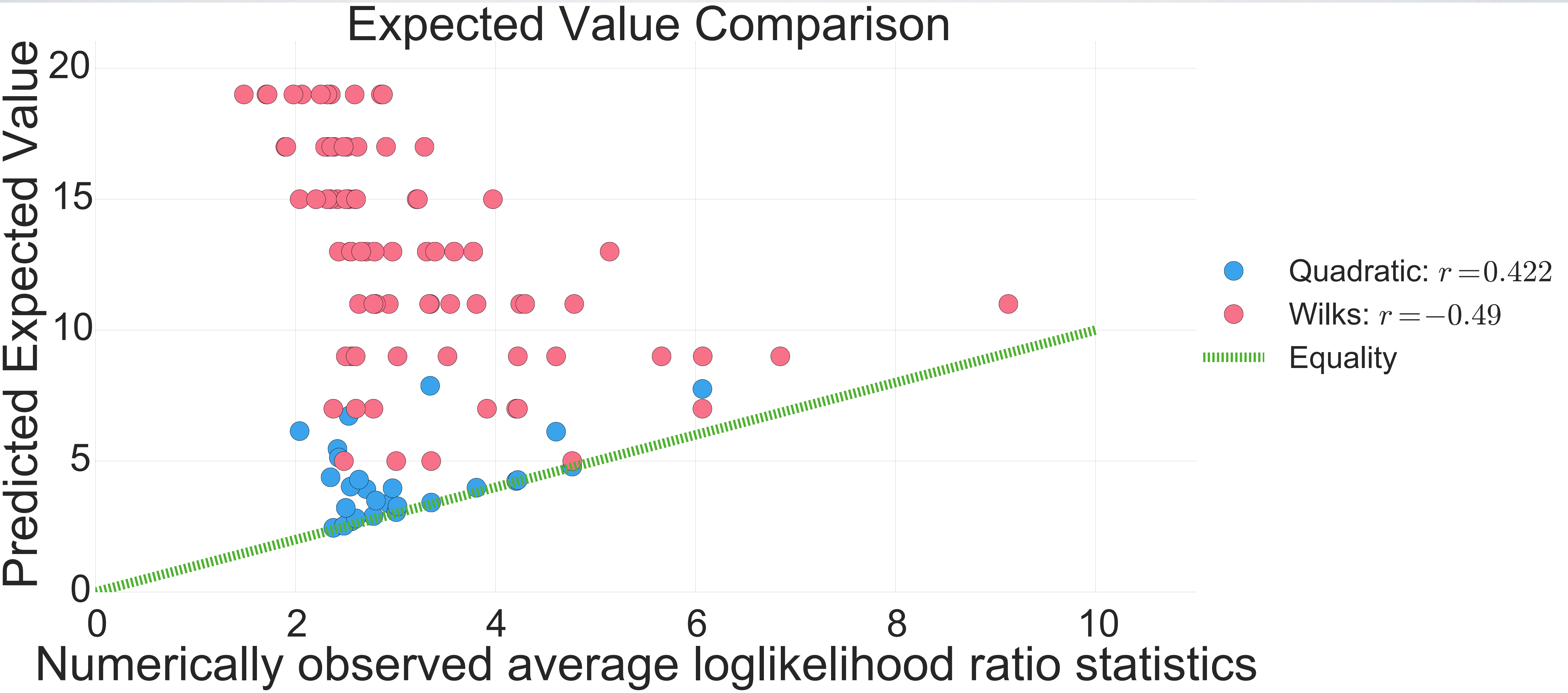
$$\langle \lambda(\rho_{true}, \mathcal{M}_d) \rangle \approx -tr\left(\left\langle \left.\frac{\partial^2 \mathcal{L}}{\partial \rho^2}\right|_{\hat{\rho}} |\rho_{true} - \hat{\rho}\rangle\rangle\langle\langle\rho_{true} - \hat{\rho}| \right\rangle \right)$$

Curvature

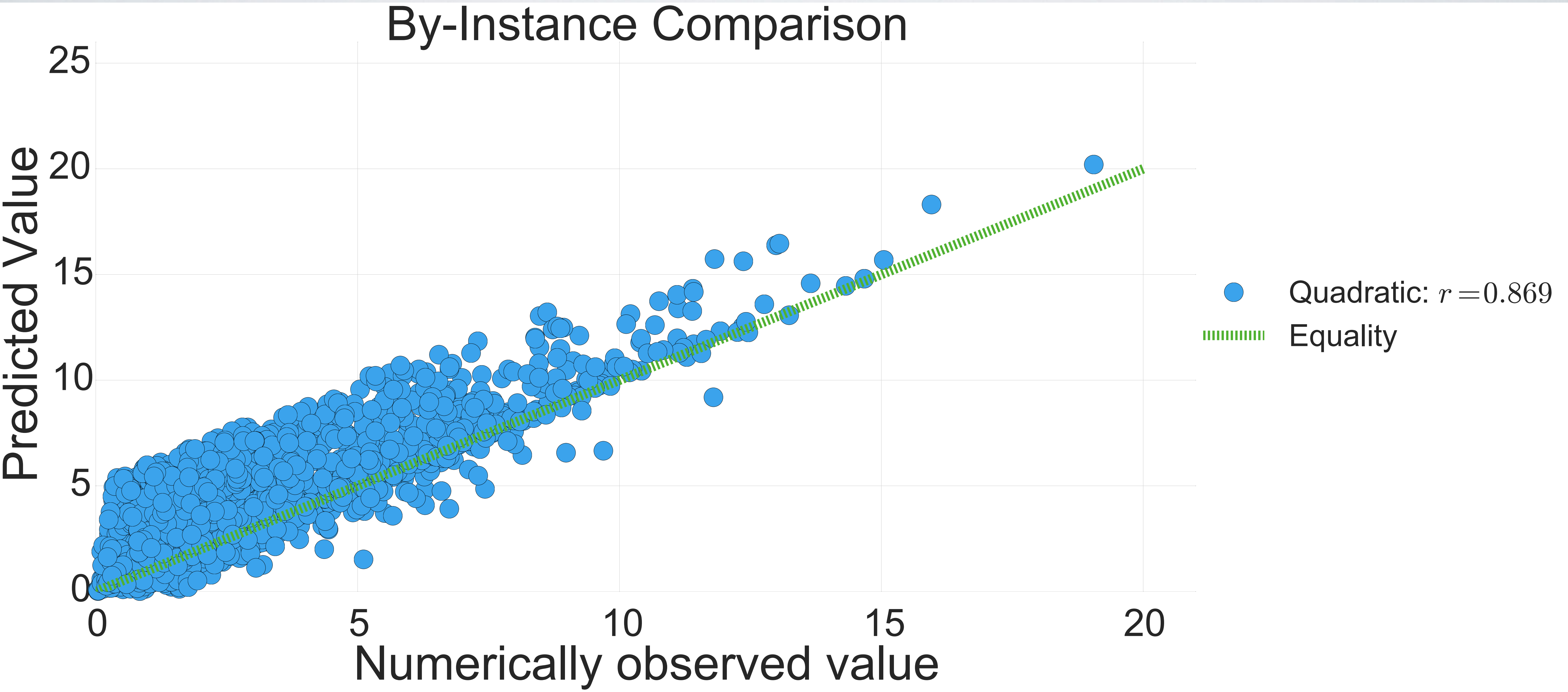
Error

Addresses the question "If estimate *close to truth*, what does the LLRS do?"

This estimator outperforms the Wilks Theorem in predicting the Monte Carlo averages.



It also performs reasonably well on an instance-by-instance basis.



In the absence of any boundaries, the prediction *reduces* to that of Wilks.

No boundaries = no bias in MLE

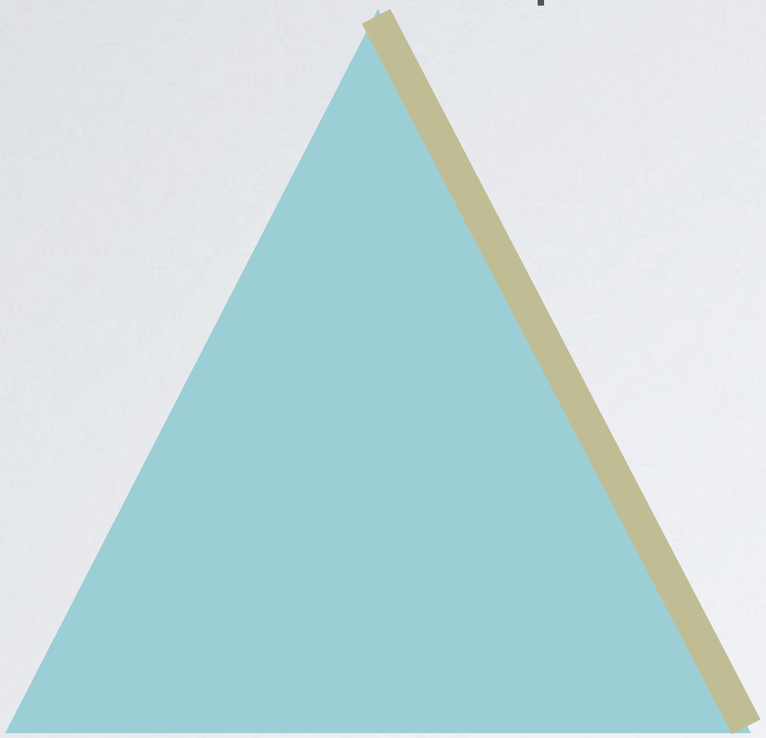
$$\langle \lambda(\rho_0, \hat{\rho}) \rangle \approx -tr \left(\left\langle \left. \frac{\partial^2 \mathcal{L}}{\partial \rho^2} \right|_{\hat{\rho}} |\rho_0 - \hat{\rho}\rangle\rangle\langle\langle\rho_0 - \hat{\rho}| \right\rangle \right)$$

$$\langle \widehat{\lambda(\rho_0, \hat{\rho})} \rangle \approx tr \left(\hat{I}(\hat{\rho}) Cov(\hat{\rho}) \right) \approx d^2 - 1$$

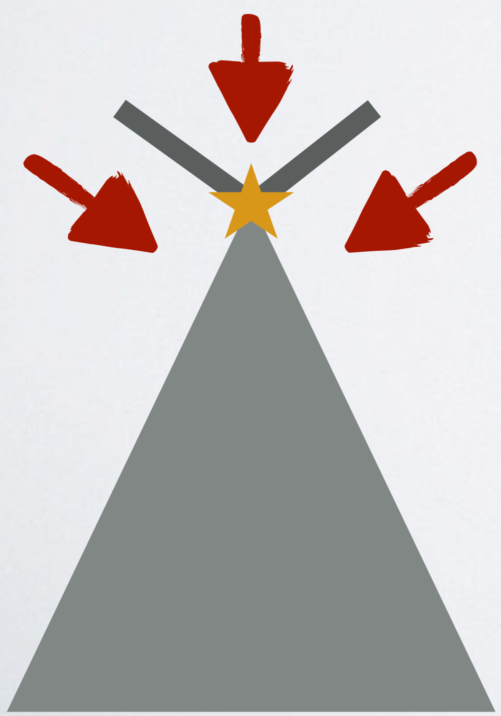
Does not predict distribution, however

What was the **point** of all that **math**?

We are going to have a way to do nested model selection in quantum tomography!

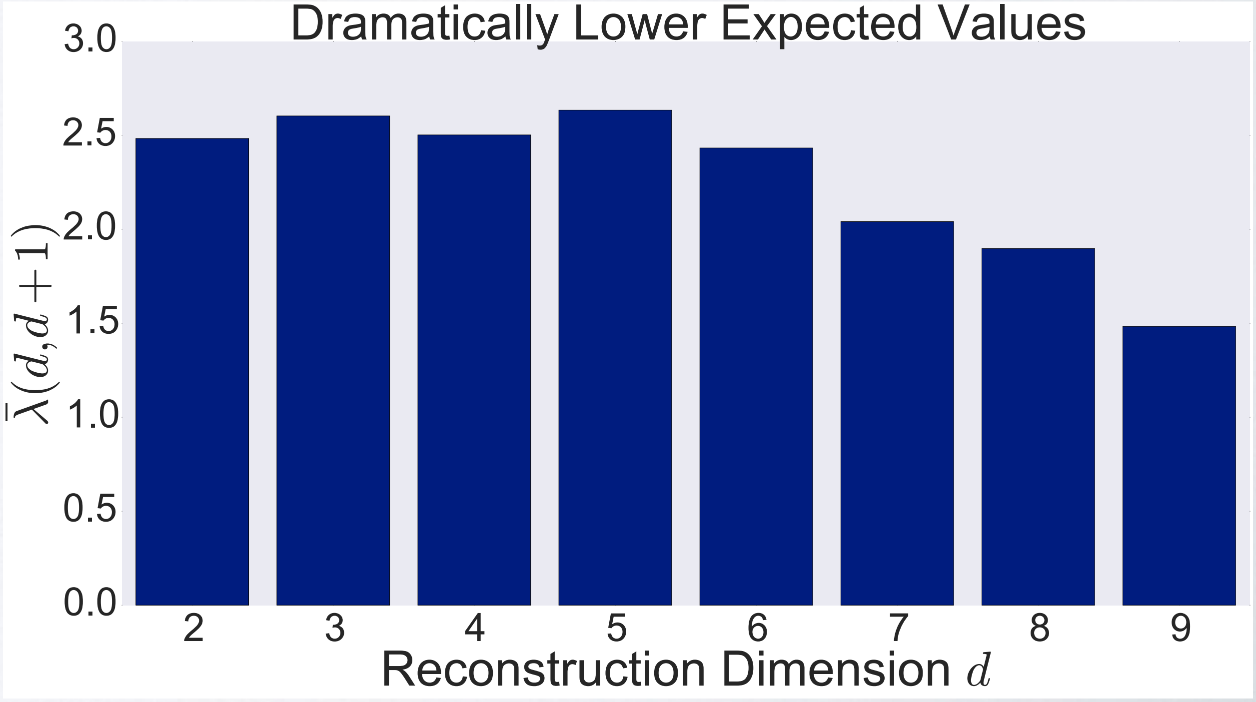


Simplex-Based Method



$$\lambda(\mathcal{M}_d, \mathcal{M}_{d+1})$$

Cannot use Wilks Theorem



What is **next**?

There are many ways forward.

Use this result to ***create estimator*** of expected value

Make a **quantum information criterion**

What's with **compressed sensing** and model selection?

A model selection rule for **displaced/squeezed states**

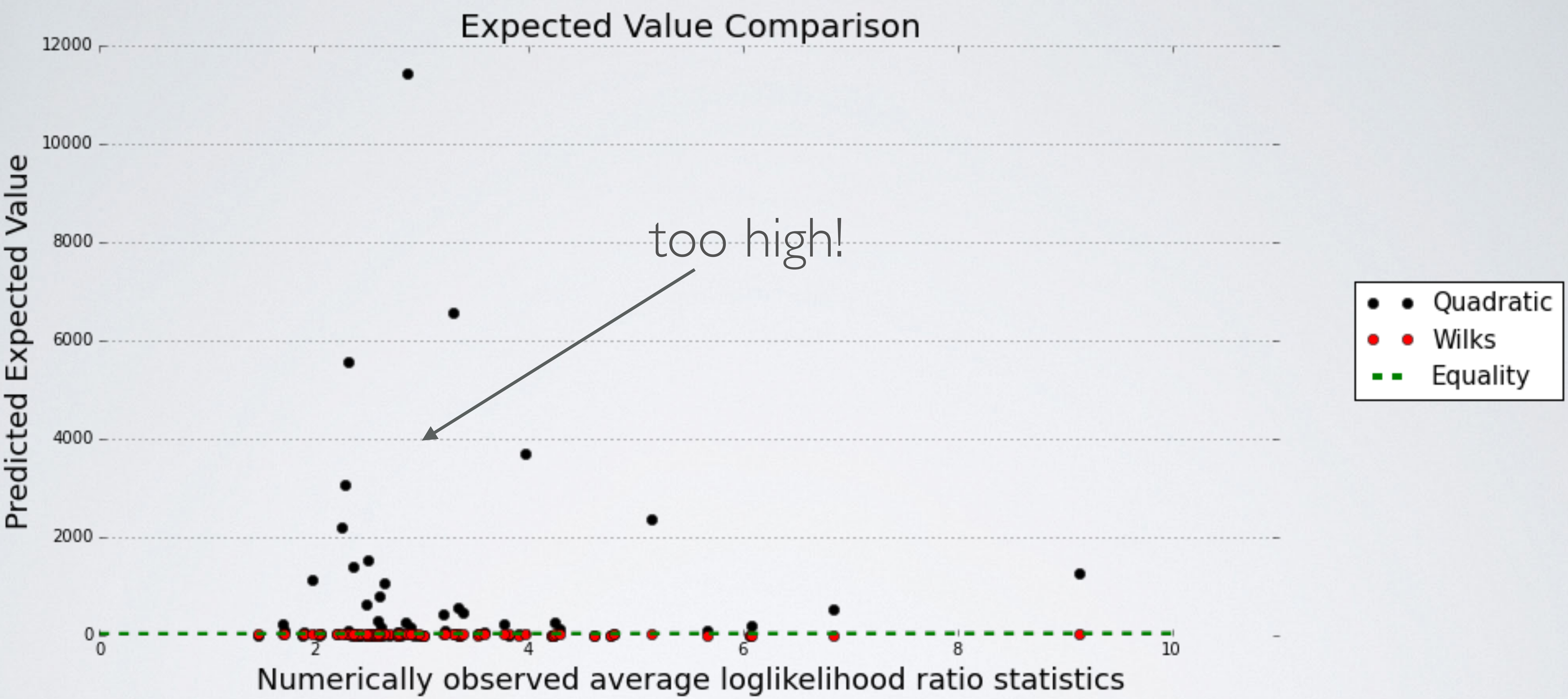
Apply active subspace methods to **speed up optimization**

A **year ago**, I thought model selection using the LLRS was **easy**…

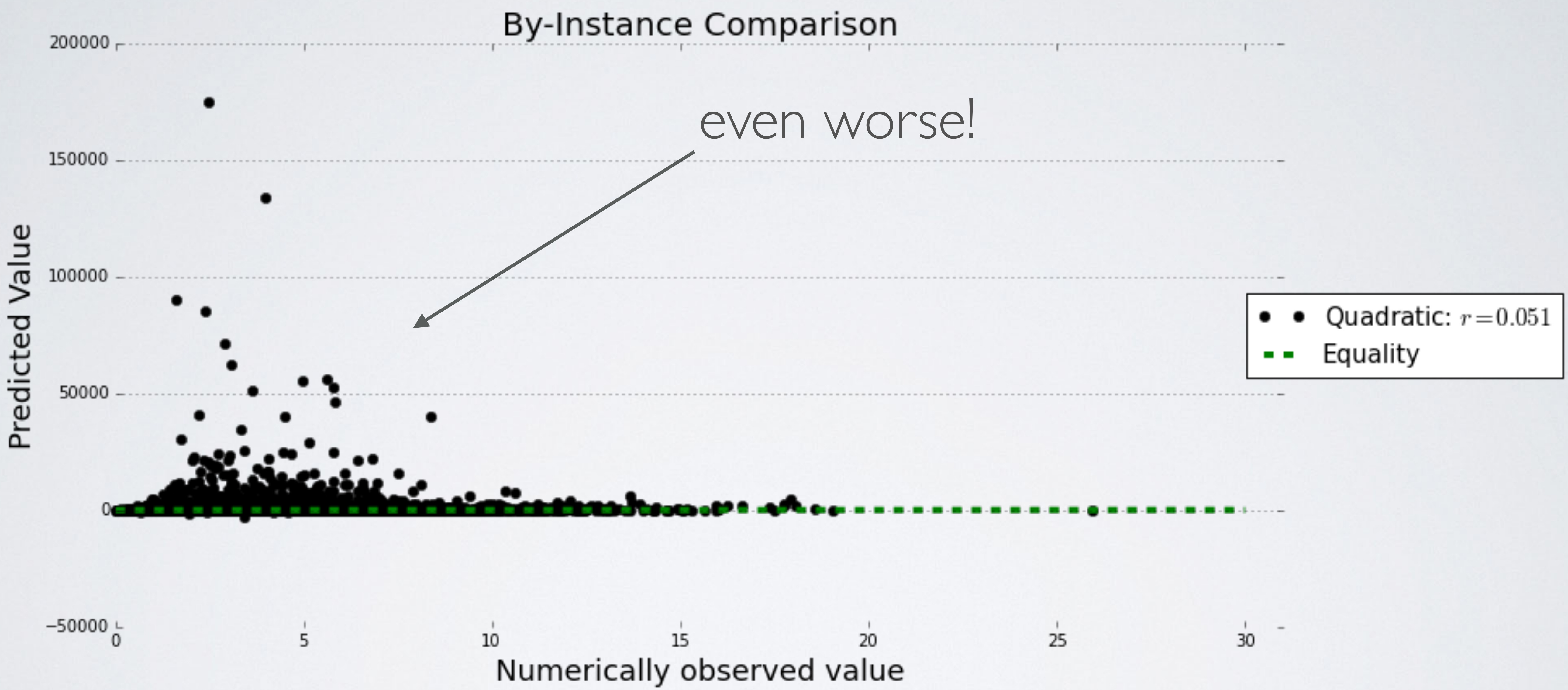
Today, I am certain it is **vastly harder** than I (and others!) thought.

Model selection in quantum state tomography **is hard** because we have to deal with **boundaries**.

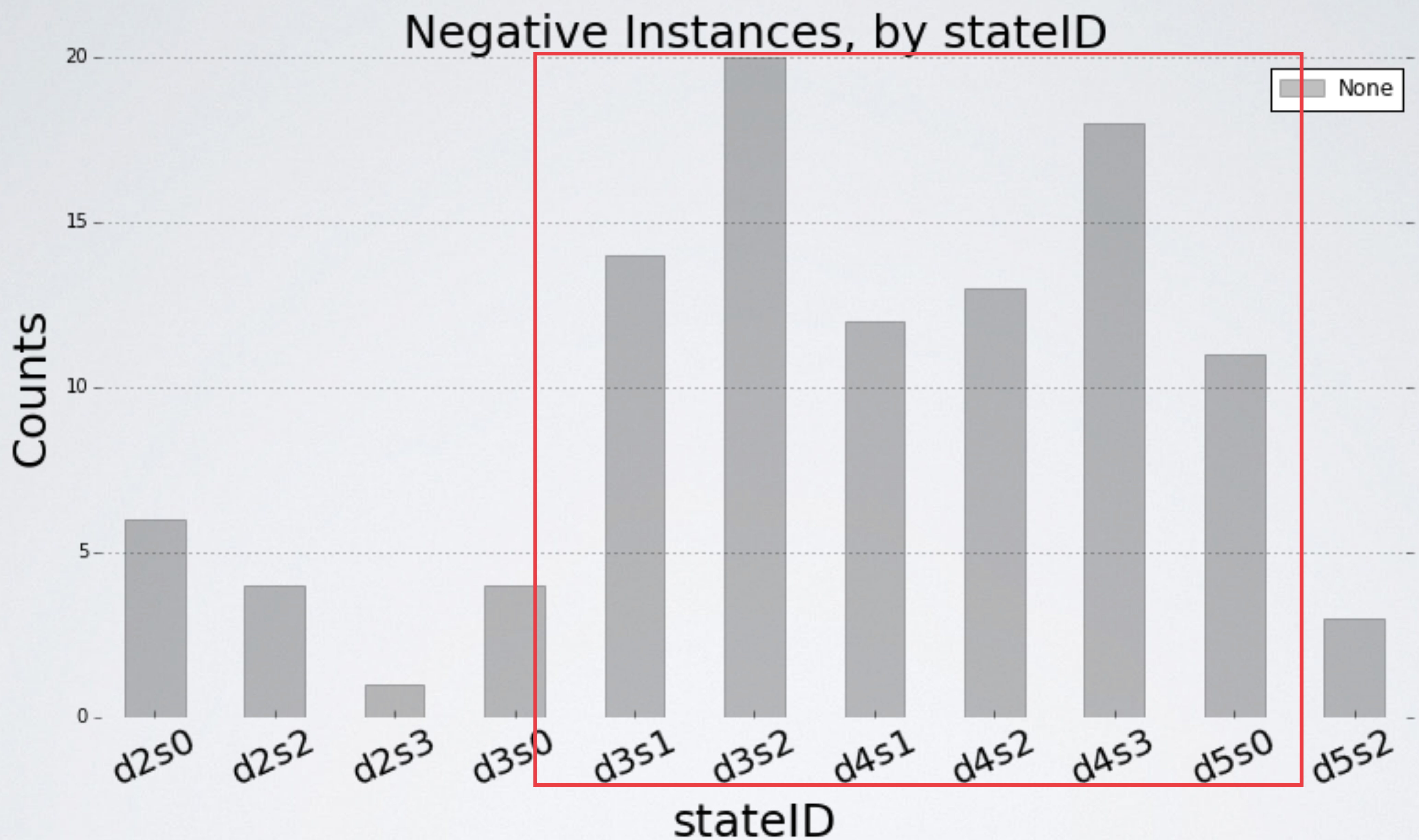
There are some problems, though:



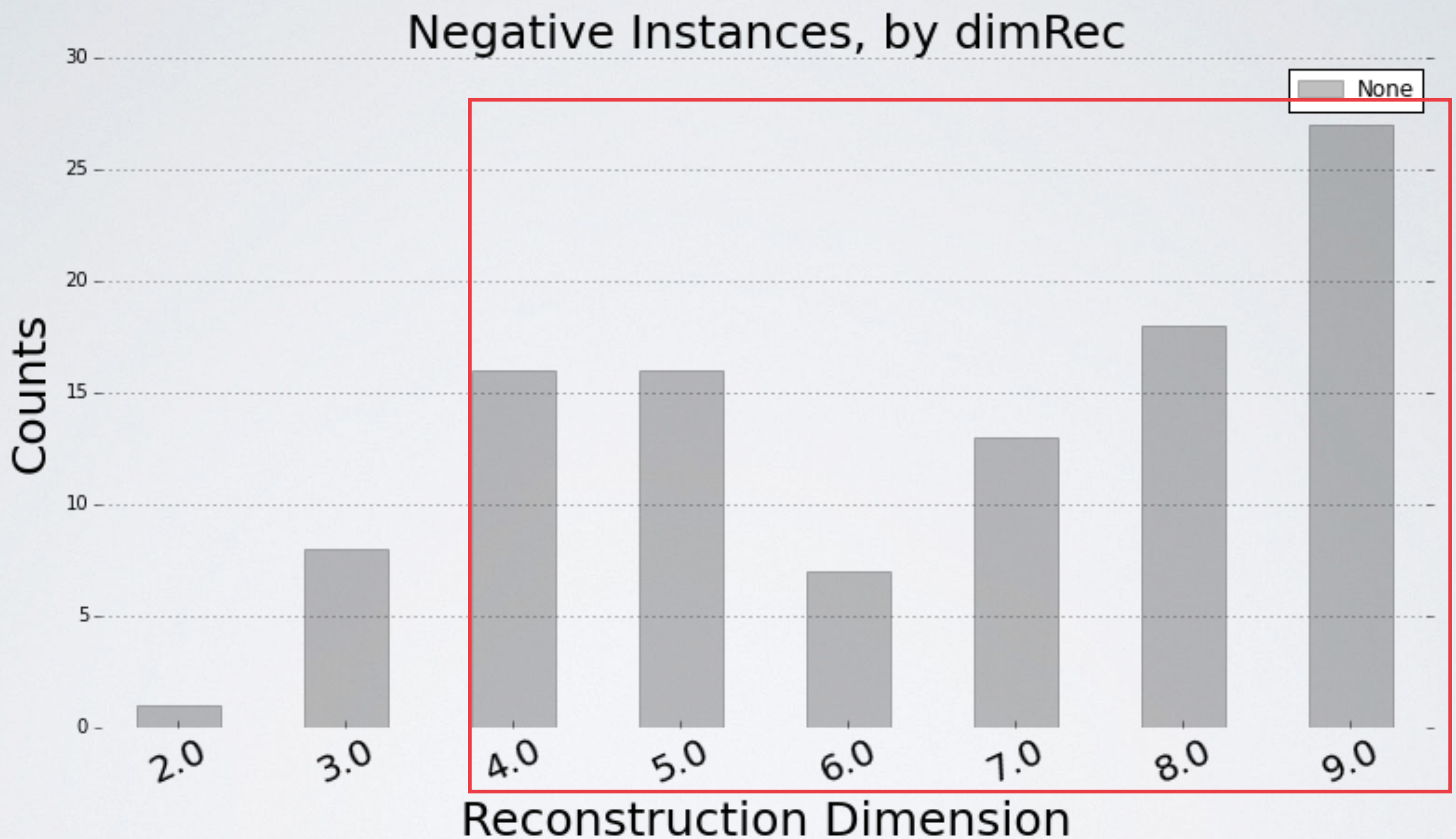
There are some problems, though:



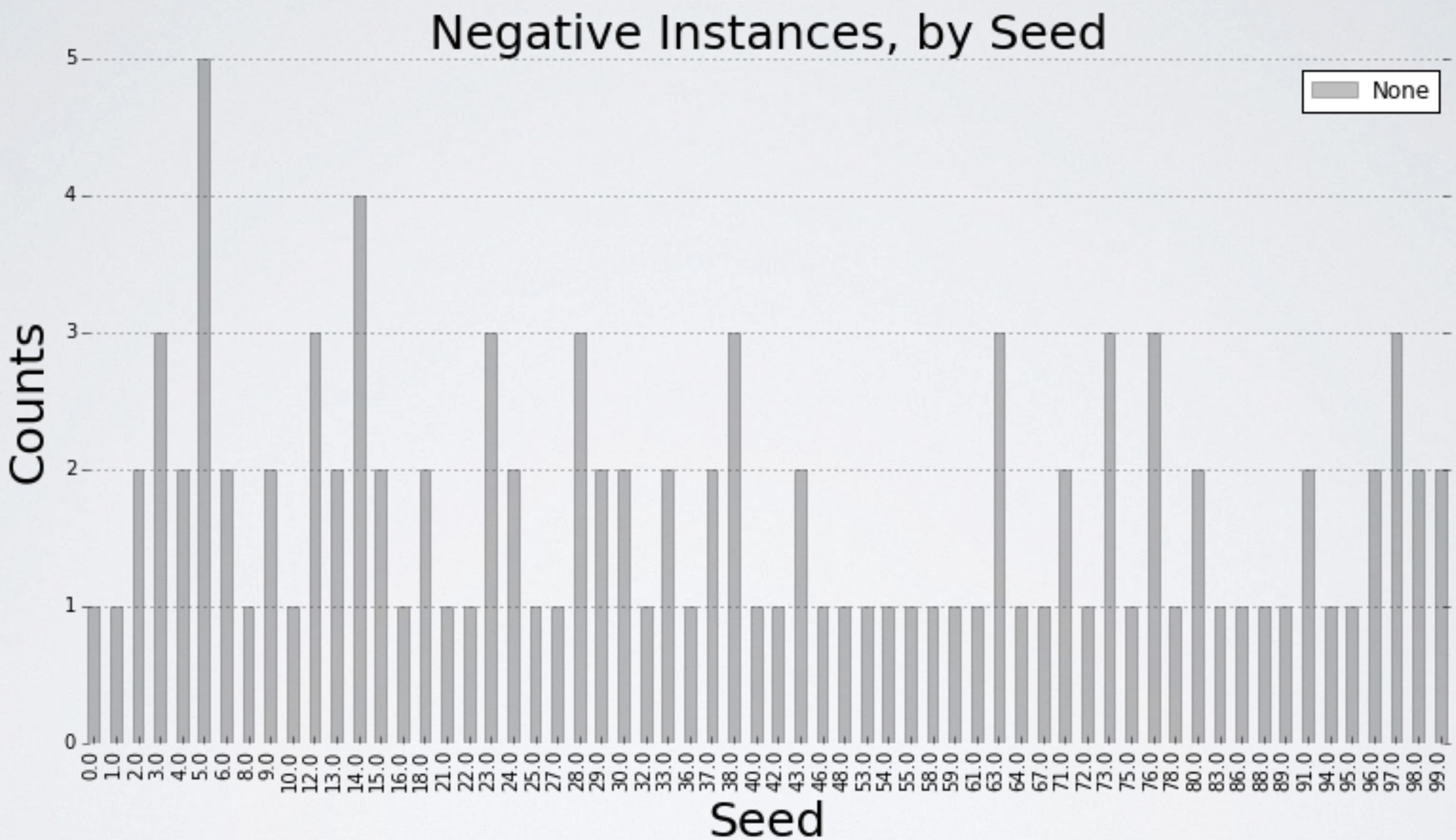
Let’s try to isolate these problem instances.



Let's try to isolate these problem instances.



Let’s try to isolate these problem instances.



Conclusions:

Little dependence on random number seed (rules out systematic issue).

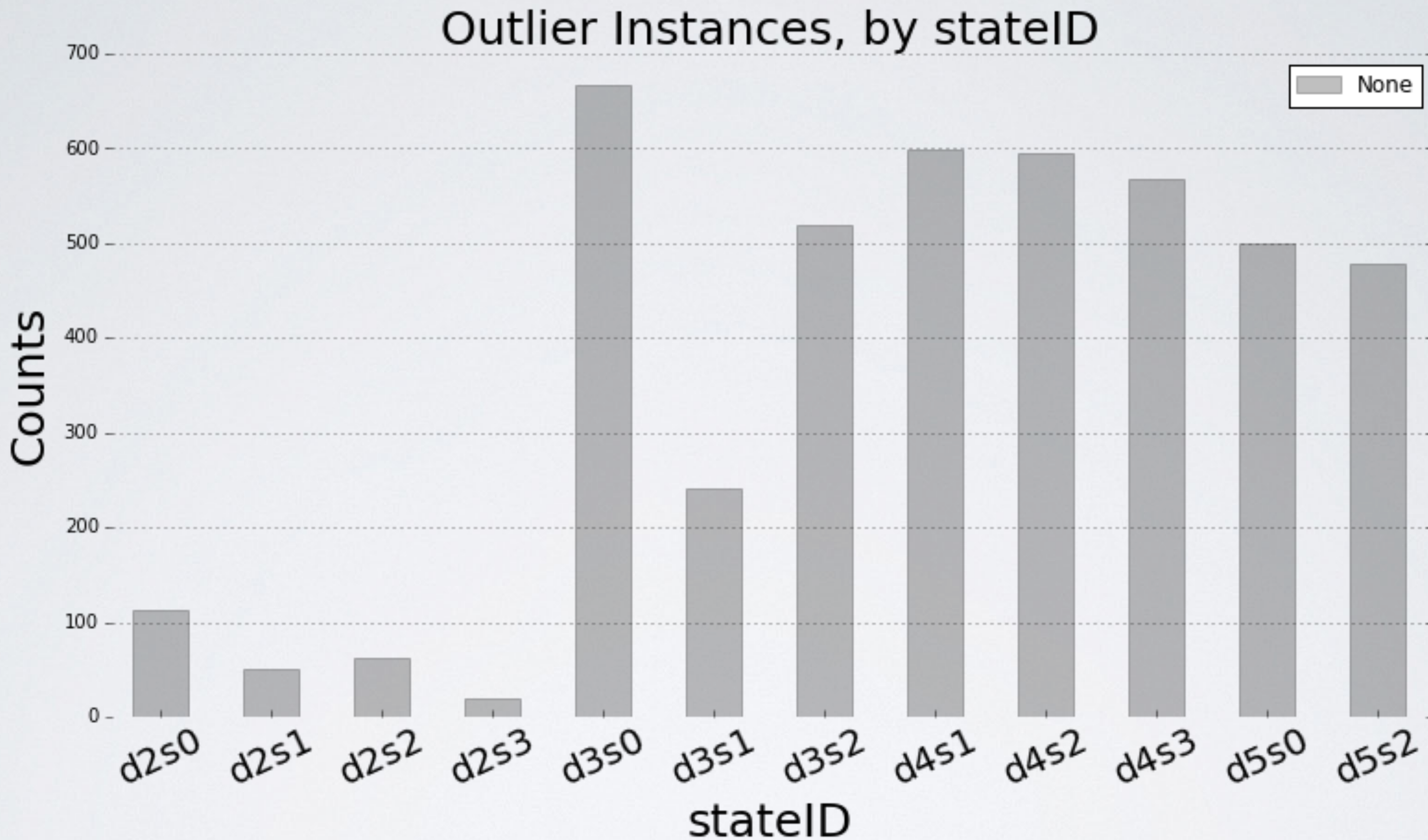
Strong dependence on stateID and dimRec

Look at **maximum sample size:**

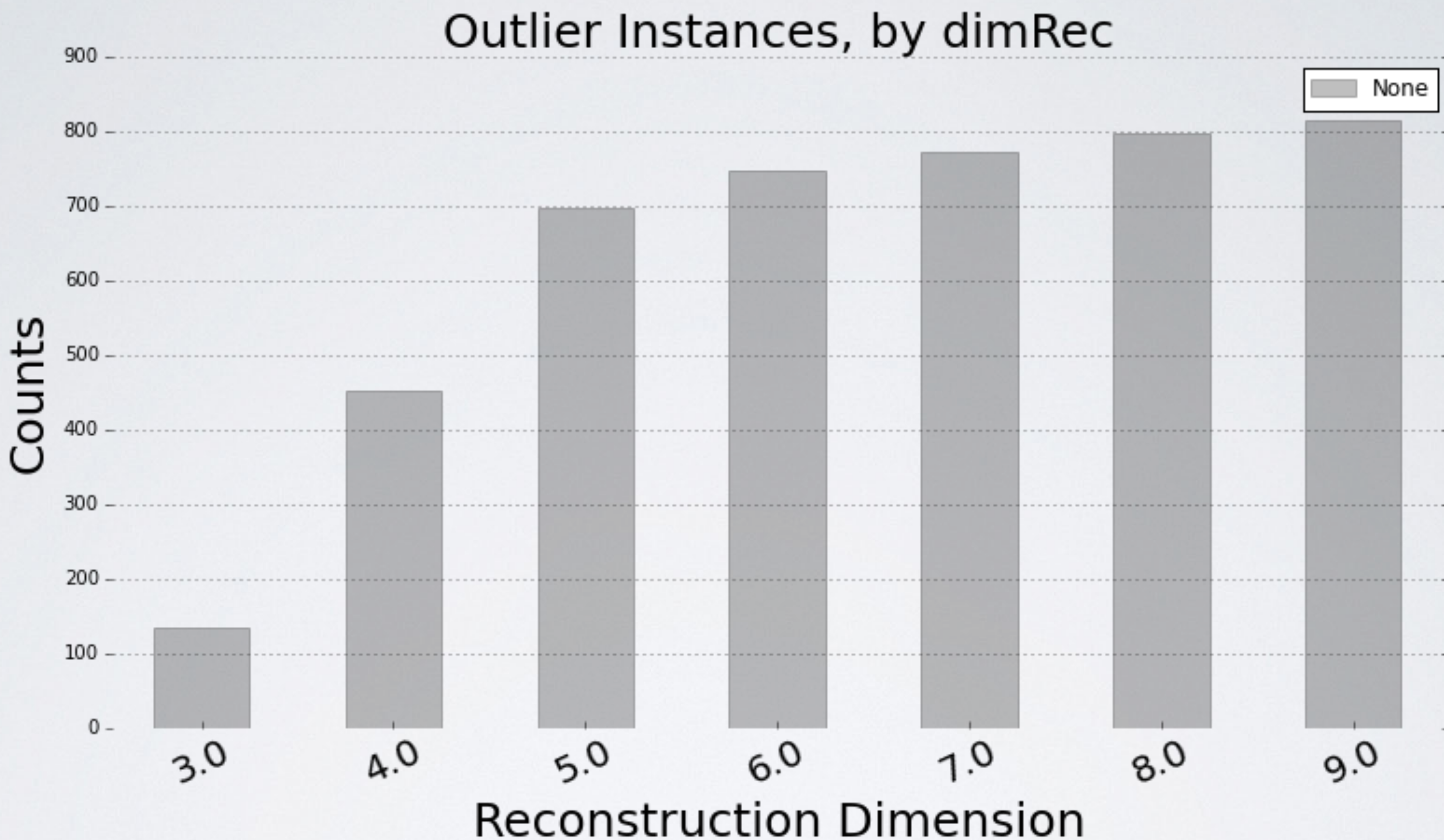
Small sample size= inaccurate estimation, particularly in high dimensions

```
[('d2s0', 49729),
 ('d2s2', 99856),
 ('d2s3', 74529),
 ('d3s0', 99856),
 ('d3s1', 10000),
 ('d3s2', 10000),
 ('d4s1', 10000),
 ('d4s2', 10000),
 ('d4s3', 10000),
 ('d5s0', 24964),
 ('d5s2', 49729)]
```

Similar results for *outliers* (predictions > 5 away from numerics).



Similar results for *outliers* (predictions > 5 away from numerics).



Similar results for *outliers* (predictions > 5 away from numerics).