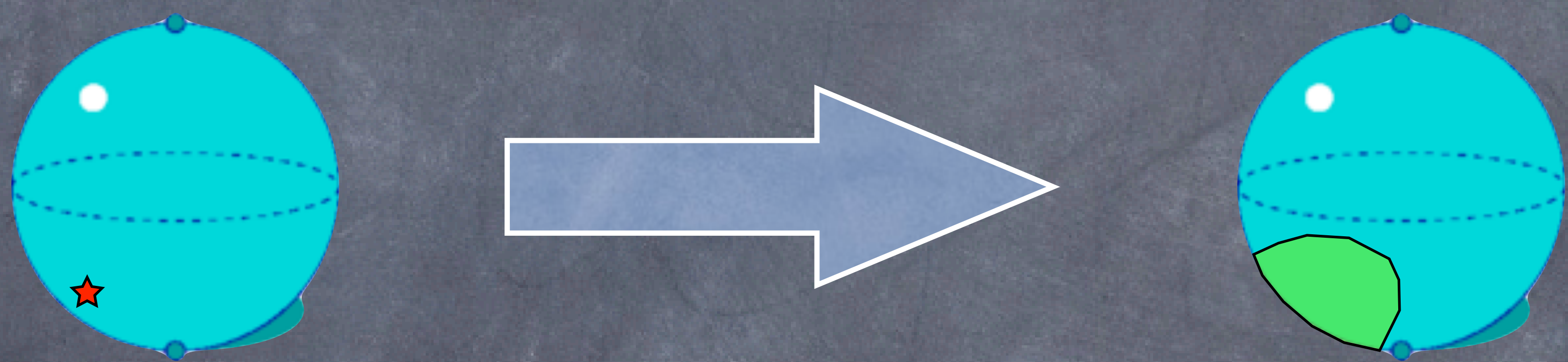
Tomography for Fault Tolerance:

Confidence regions for quantum hardware



Robin Blume-Kohout

CQIQC-IV

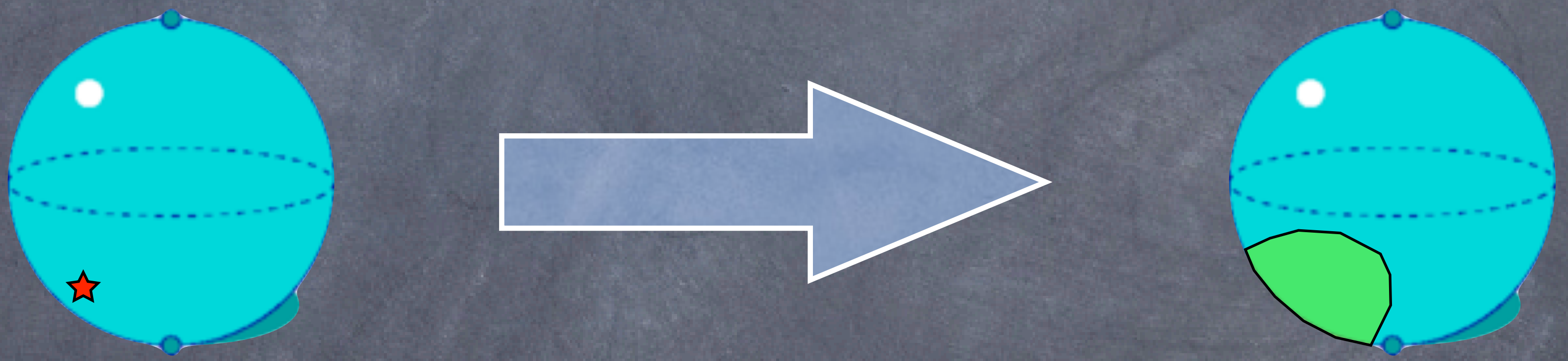
Aug. 10, 2011

Los Alamos NATIONAL LABORATORY EST. 1943

UNM

My talk in a nutshell

- When we do tomography, we should really report a *region* -- not a *point* -- in state (process) space.



- Here's how.

A larger nutshell

1. We use tomography to characterize quantum states and quantum gates.

2. The ultimate application is a fault-tolerant quantum computer.
   Our tomography must be *very* accurate, precise, and reliable.

3. Current methods -- *point estimators* -- are not up to the task. The solution is *region estimators*, which support rigorous conclusions about devices.

4. *Confidence region* estimators are (currently) the best approach.

5. The most powerful ones are *probability ratio* (PR) estimators.

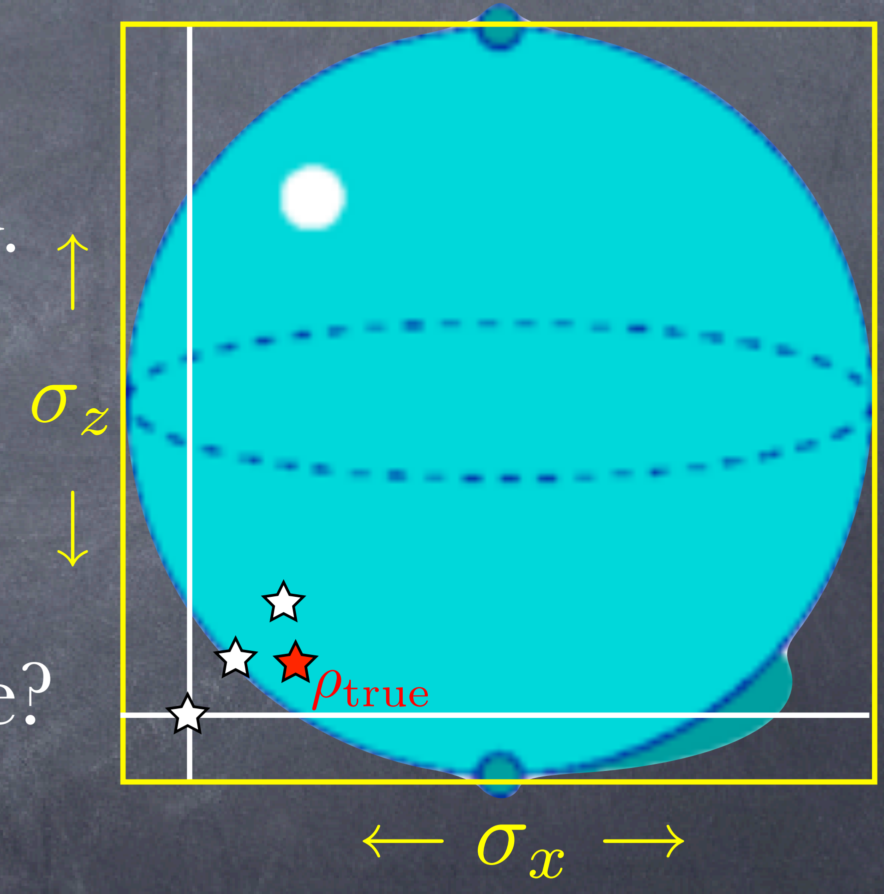
6. Among PR estimators, *likelihood ratio* (LR) estimators are:
   (1) Simple, elegant, and easy to implement.
   (2) Nearly optimal in the sense of worst-case performance.

7. A complete algorithm for computing LR regions...
   ...but an important mathematical ingredient could be <u>improved</u>.

Tomography

- We want to estimate a $d \times d$ density matrix ρ.
  (*process tomography* is isomorphic, so I won't discuss it explicitly)

  1. Make measurements with at least $d^2 - 1$ linearly independent outcomes $\{M_1 \ldots M_K\}$.
  2. Use the observed frequencies $\{n_k\}$ to estimate ρ.



- What *estimator* -- $\hat{\rho}\left(\{n_k\}\right)$ -- should we use?

  1. Linear inversion. *Yields negative estimates.*
  2. Maximum likelihood. *Yields rank-deficient estimates.*
  3. Bayesian inference. *Hard. Requires choosing a prior.*
  4. L_1 regularization (compressed sensing). *Low-rank bias.*

The Central Problem

- Need "error bars" on any point estimate $\hat{\rho}$!

  Why?

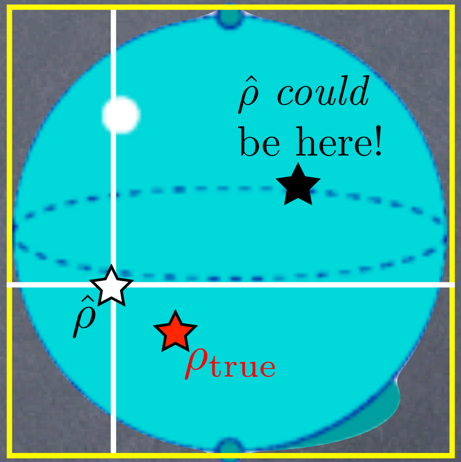
- Because $\hat{\rho}$ is never exactly the true state.
  $\Rightarrow$ At best, it's close. Probably.
  $\Rightarrow$ By itself, $\hat{\rho}$ has no meaning!

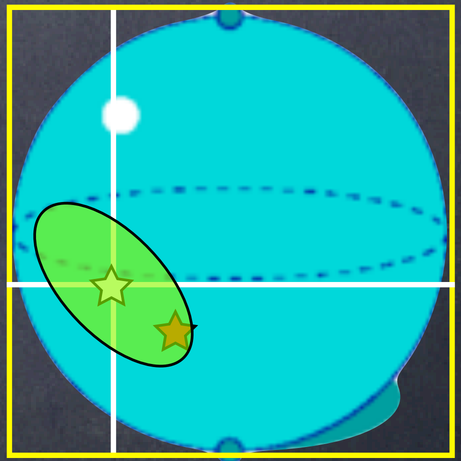
- Quantifying how close and how probably converts a **point estimator** $\hat{\rho}$ into a **region estimator** $\hat{\mathcal{R}}$.

- What kind of region estimator do we want?

  $\Uparrow\Downarrow$

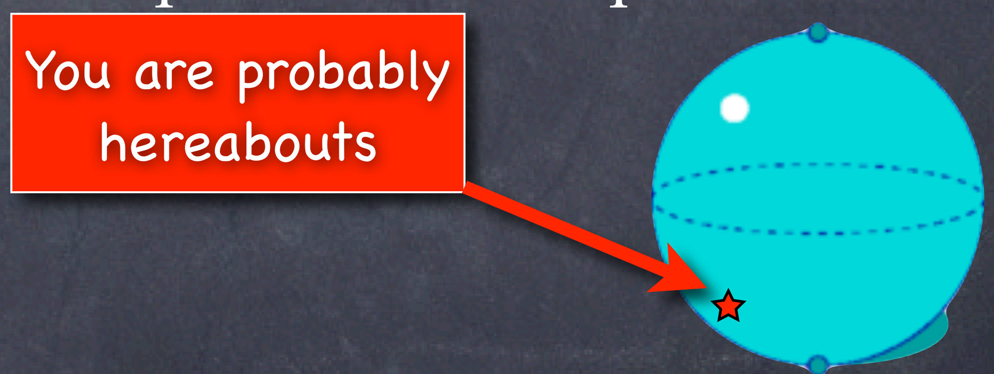
  What are we going to do with the estimate?





Application: Fault Tolerance

- Noise is deadly to [quantum] computations
  ⇒ we need 10^3 - 10^{22} *consecutive* perfect gates!
  ⇒ quantum gates can fail in weird ways (e.g., superposition)
- So we design FT architectures robust to *specific* errors
  ⇒ nested error correction schemes
  to simulate 1 good gate using N noisy gates.
- Phase transition at the *threshold error probability* (10^{-8} - 10^{-3}):
  ⇒ *below* the threshold, FT design works as $N \rightarrow \infty$.
  ⇒ *above* the threshold, reliable computation is impossible.
- Requires very accurate and ***very*** reliable tomography.



Application: Fault Tolerance

- Noise is deadly to [quantum] computations
  ⇒ we need 10^3 - 10^{22} *consecutive* perfect gates!
  ⇒ quantum gates can fail in weird ways (e.g., superposition)

- So we design FT architectures robust to *specific* errors
  ⇒ nested error correction schemes
  to simulate 1 good gate using N noisy gates.

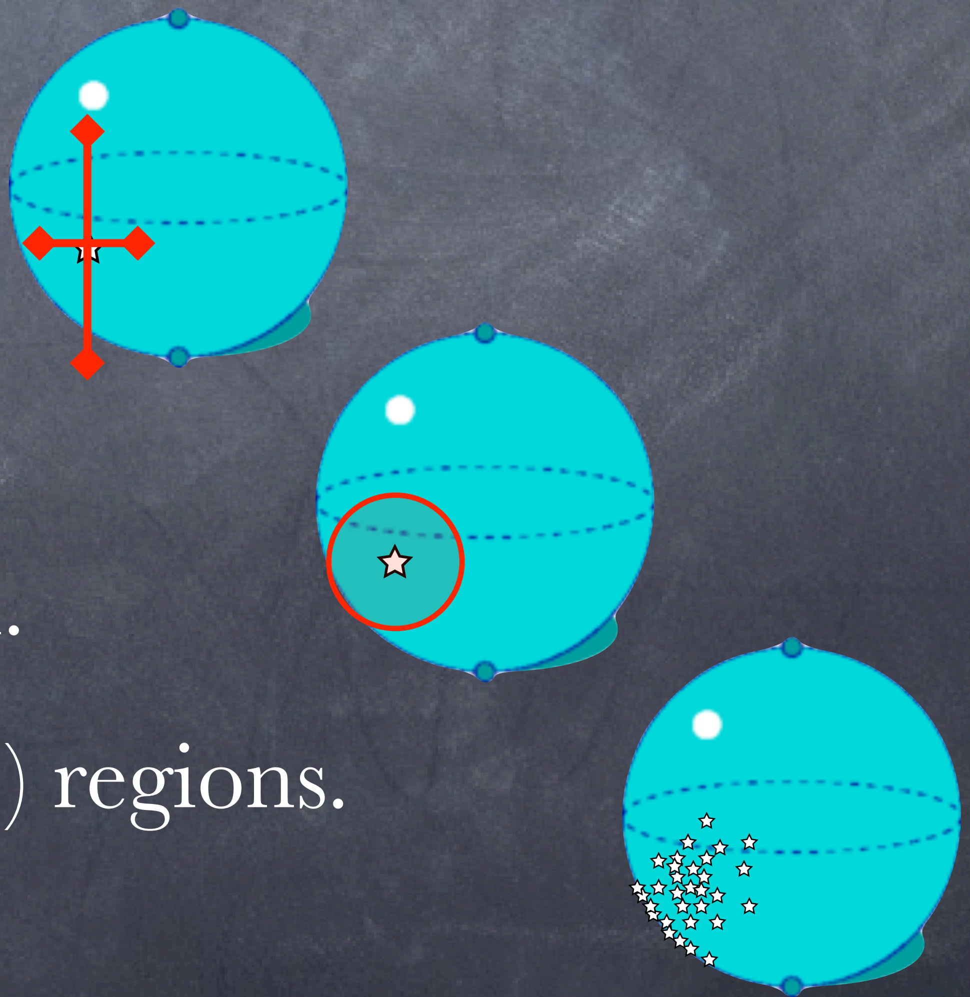
- Phase transition at the *threshold error probability* (10^{-8} - 10^{-3}):
  ⇒ *below* the threshold, FT design works as $N\to\infty$.
  ⇒ *above* the threshold, reliable computation is impossible.

- Requires very accurate and ***very*** reliable tomography.

Solution: Region Estimators

- Point estimators map data ⇒ a single state.
- Region estimators map data ⇒ a region of states
- Some ad-hoc examples:

1. A point estimate
   + error bars.
2. A point estimate
   + large deviation bound.
3. Bootstrap ("resampling") regions.



Why Region Estimators?

- Region estimators can have a rigorous meaning:

  "With probability at least α, the true state is in $\hat{\mathcal{R}}$!"*

- Region estimators can justify statements like:

  "My ancillae satisfy the fault tolerance threshold with 99.999999% certainty."*

- Point estimators **cannot** do these things.

* Stay tuned for precise interpretation & clarification!

So what's a region estimator?

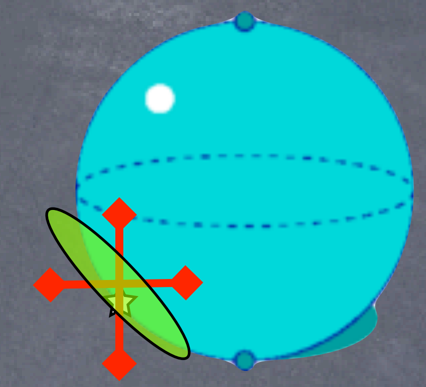
- A map $\hat{\mathcal{R}}$, from data $\Rightarrow$ sets of states
- If you observe data D, you report $\hat{\mathcal{R}}(D)$.
- **Desiderata**:

CORRECTNESS
- $\hat{\mathcal{R}}(D)$ should contain ρ with probability α.

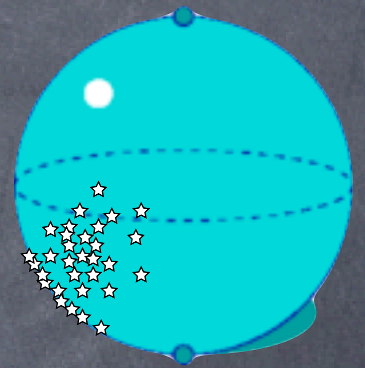
POWER
- $\hat{\mathcal{R}}(D)$ should be as small as possible.

(convenience)
- $\hat{\mathcal{R}}(D)$ should be connected & convex.

FAQ



- "Can't we just put error bars on each measurement?"
  No. There are usually <u>co</u>-variances, too.
  And your error regions would go outside the Bloch sphere.
  And it doesn't work anyway because of the Bonferroni effect.



- "Can't we just use MLE and do bootstrapping?"
  No. Bootstrap isn't reliable for biased estimators.

- "Okay. But, our 'error bars' will be centered around the tomographic point estimate, right?"
  No. That's like putting wings on a car, instead of designing an airplane. We'll derive optimal regions from scratch.

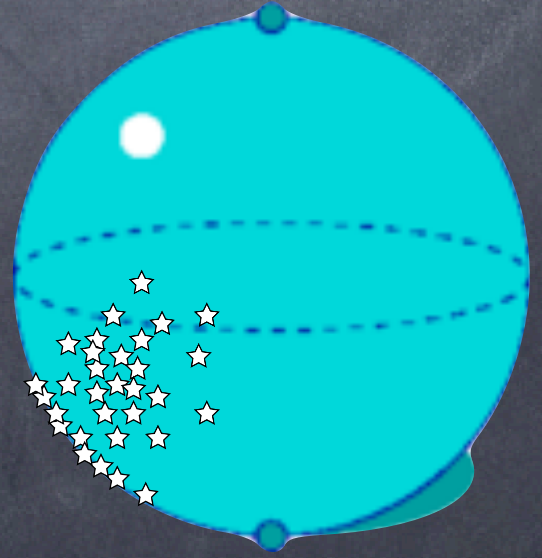
- "Can't you just tell us the answer!"
  Okay. But I need a couple more slides first...

Why not Bootstrap?

- The bootstrap calculates standard errors
  $=$ variance of a point estimator $\hat{\rho}$ (e.g., $\hat{\rho}_{\mathrm{MLE}}$)

- "Resampling": use $\hat{\rho}_{\mathrm{MLE}}$ to generate many fake datasets. Do MLE on them. Look at the variance of the cloud of MLE estimates.

- Unreliable! because:
  1. MLE is not unbiased.
  2. Bootstrap calculates standard error at $\hat{\rho}_{\mathrm{MLE}}$ <u>not</u> at the true ρ!



- Standard errors are no good.

Why not Bootstrap?

- The bootstrap calculates standard errors
  = variance of a point estimator $\hat{\rho}$ (e.g., $\hat{\rho}_{\mathrm{MLE}}$)

- "Resampling": use $\hat{\rho}_{\mathrm{MLE}}$ to generate many fake datasets. Do MLE on them. Look at the variance of the cloud of MLE estimates.

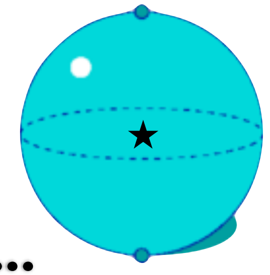
- Unreliable! because:
  1. MLE is not unbiased.
  2. Bootstrap calculates standard error at $\hat{\rho}_{\mathrm{MLE}}$ <u>not</u> at the true ρ!

- Standard errors are no good.

Example 1: Biased estimators bad!

Consider the <u>stupid</u> estimator:

$$\hat{\rho}(\text{data}) = \frac{1}{2}\mathbb{1}$$



Its variance is always zero...
...so the bootstrap says "I have no uncertainty about ρ at all!" ☹

Why not Bootstrap?

- The bootstrap calculates standard errors
  = variance of a point estimator $\hat{\rho}$ (e.g., $\hat{\rho}_{\mathrm{MLE}}$)

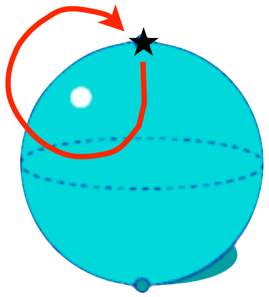
- "Resampling": use $\hat{\rho}_{\mathrm{MLE}}$ to generate many fake datasets. Do MLE on them. Look at the variance of the cloud of MLE estimates.

- Unreliable! because:
  1. MLE is not unbiased.
  2. Bootstrap calculates standard error at $\hat{\rho}_{\mathrm{MLE}}$ <u>not</u> at the true ρ!

- Standard errors are no good.

Example 2: Circular reasoning

Let's do MLE on a qubit.
Measure Z once. Get "up".



$$\hat{\rho}_{\mathrm{MLE}} = |{\uparrow}\rangle\langle{\uparrow}|$$

$\Rightarrow$ <u>every</u> simulated dataset is "up"

$\Rightarrow$ <u>every</u> estimate is $\hat{\rho} = |{\uparrow}\rangle\langle{\uparrow}|$

$\Rightarrow$ standard errors vanish!

Good Region Estimators

Desideratum: $\hat{\mathcal{R}}(D)$ should contain ρ w/probability α.

Good Region Estimators

Desideratum: $\hat{\mathcal{R}}(D)$ should contain ρ w/probability α.

- Credible Region Estimator
- Bayesian. Requires a prior $Pr(\rho)$.
- Uses data to calculate

$$Pr(\rho|\text{data})$$

- Assigns $\hat{\mathcal{R}}(D)$ satisfying

$$Pr\left(\rho \in \hat{\mathcal{R}}(D)|D\right) \geq \alpha$$

- GOOD: nice interpretation, easy to understand.
- BAD: only valid if rho is really drawn from $Pr(\rho)$.

Good Region Estimators

Desideratum: $\hat{\mathcal{R}}(D)$ should contain ρ w/probability α.

- Credible Region Estimator
- Bayesian. Requires a prior $Pr(\rho)$.
- Uses data to calculate

$$Pr(\rho|\text{data})$$

- Assigns $\hat{\mathcal{R}}(D)$ satisfying

$$Pr\left(\rho \in \hat{\mathcal{R}}(D)|D\right) \geq \alpha$$

- GOOD: nice interpretation, easy to understand.
- BAD: only valid if rho is really drawn from $Pr(\rho)$.

- Confidence Region Estimator
- Non-Bayesian. No prior!
- Must satisfy:

$$Pr\left(\hat{\mathcal{R}}(D) \ni \rho|\rho\right) \geq \alpha \quad \ldots \forall\ \rho$$

- ...which implies, unconditionally

$$Pr\left(\hat{\mathcal{R}}(D) \ni \rho\right) \geq \alpha$$

- GOOD: absolutely reliable, even if your worst enemy chooses ρ!
- BAD: interpretation is less straightforward & intuitive.

Good Region Estimators

Desideratum: $\hat{\mathcal{R}}(D)$ should contain ρ w/probability α.

- Credible Region Estimator
  - Bayesian. Requires a prior $Pr(\rho)$.
  - Uses data to calculate

    $$Pr(\rho|\text{data})$$

  - Assigns $\hat{\mathcal{R}}(D)$ satisfying

    $$Pr\left(\rho \in \hat{\mathcal{R}}(D)|D\right) \geq \alpha$$

  - GOOD: nice interpretation, easy to understand.
  - BAD: only valid if rho is really drawn from $Pr(\rho)$.

- Confidence Region Estimator
  - Non-Bayesian. No prior!
  - Must satisfy:

    $$Pr\left(\hat{\mathcal{R}}(D) \ni \rho|\rho\right) \geq \alpha \ \ldots \forall\ \rho$$

  - ...which implies, unconditionally

    $$Pr\left(\hat{\mathcal{R}}(D) \ni \rho\right) \geq \alpha$$

  - GOOD: absolutely reliable, even if your worst enemy chooses ρ!
  - BAD: interpretation is less straightforward & intuitive.

Confidence Regions

- Defining property of confidence regions is *correctness*:

$\hat{\mathcal{R}}(D)$ is an α-confidence region estimator
iff
$$Pr(\hat{\mathcal{R}}(D) \ni \rho | \rho) \geq \alpha \quad \ldots \forall \; \rho$$

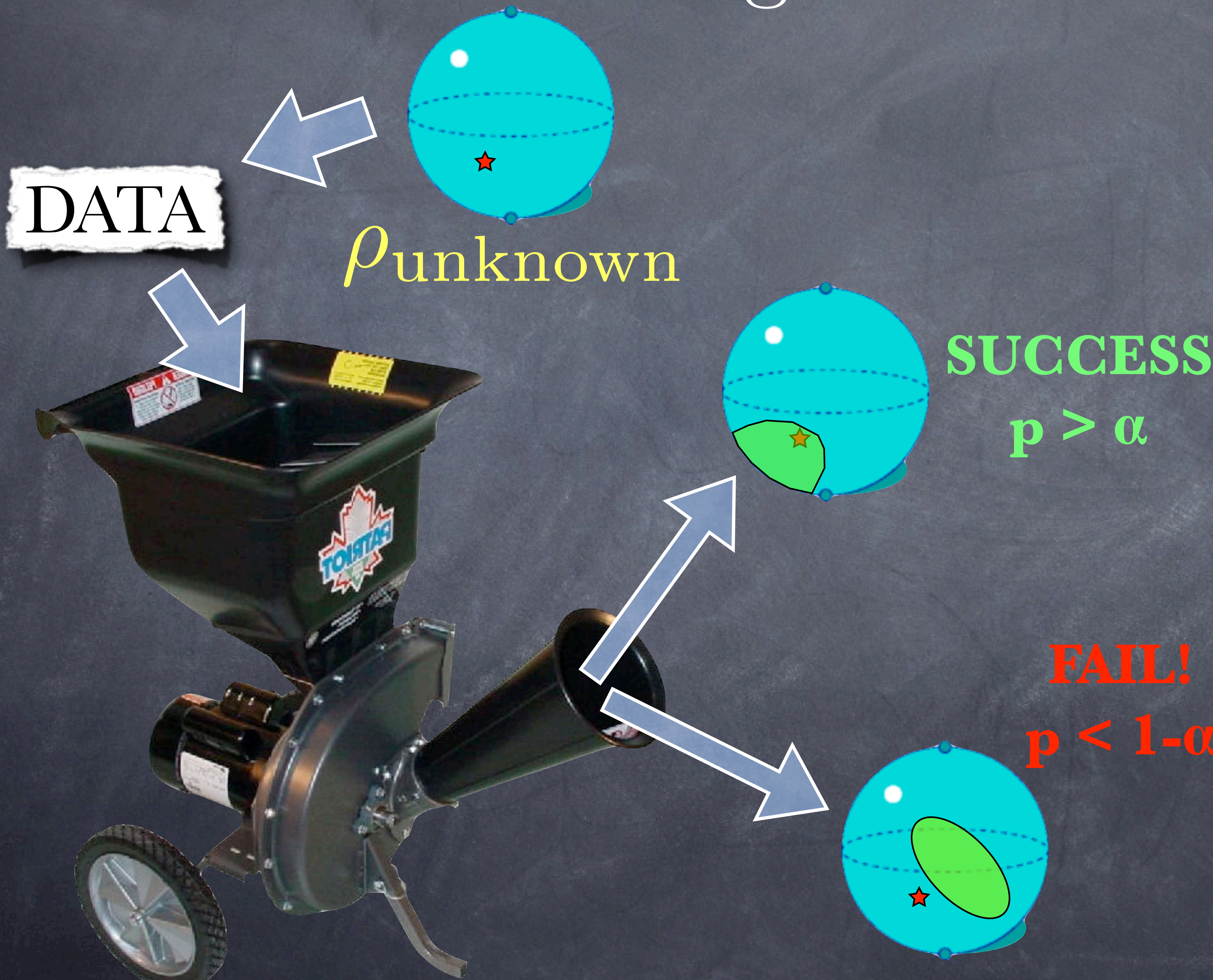
- Does NOT imply:
"Given data D, ρ is in $\hat{\mathcal{R}}(D)$
with probability α."

⟺

$Pr(\hat{\mathcal{R}}(D) \ni \rho | D) \geq \alpha$
requires a prior!

- Instead: "This procedure (estimator) will `succeed' with probability at least α".
$\Rightarrow$ "succeed" = "yield a region that contains ρ."

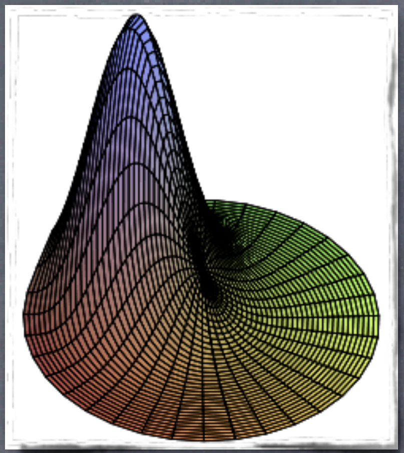
Confidence Region Estimators



1-α can easily be <u>really</u> small. e.g., 10^{-9}

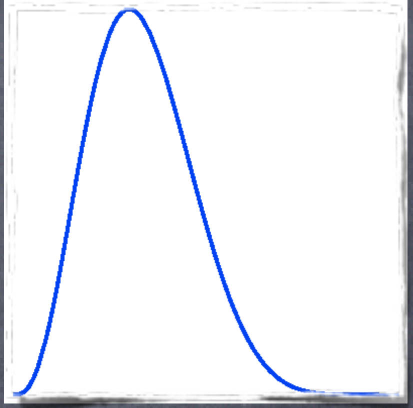
Likelihood Ratio Confidence Regions

The Likelihood Function



- Q: What do data D, generated by an unknown state ρ, tell us about ρ?
- A: Everything is contained in the likelihood function

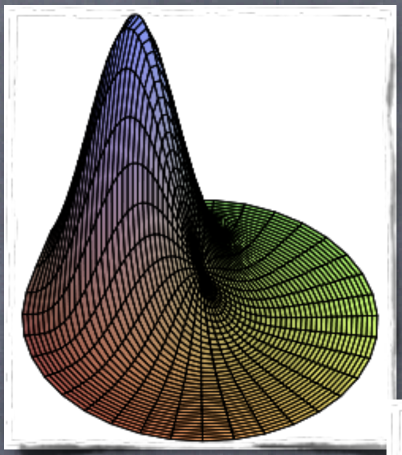
$$\mathcal{L}(\rho) \equiv Pr(D|\rho) = \prod_k \mathrm{Tr}[\mathcal{M}_k \rho]^{n_k}$$



- Properties of $\mathcal{L}(\rho)$:
  ⟹ It is *not* a probability distribution over ρ.
  ⟹ It is a *function* -- and assigns a value to each ρ.
  ⟹ It quantifies *relative* plausibility of different states.
  ⟹ Its overall magnitude has no meaning.
- The maximum of $\mathcal{L}(\rho)$ is a popular point estimator:

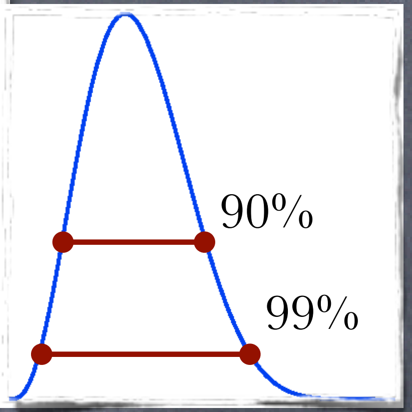
$$\hat{\rho}_{\mathrm{MLE}} = \mathrm{argmax}\,[\mathcal{L}(\rho)]$$

Recipe: LR confidence regions



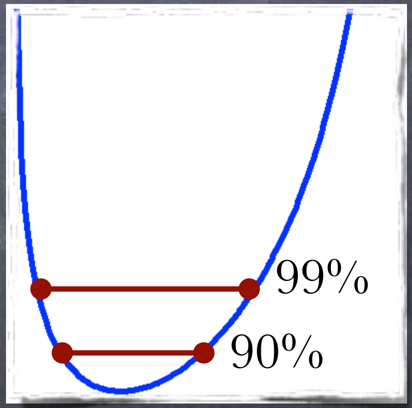
1. Take data $D = \{E_1, E_2, E_3, \ldots E_N\}$

2. Calculate the likelihood, $\mathcal{L}(\rho) = \prod_k \mathrm{Tr}(E_k \rho)^{n_k}$



3. Normalize by its maximum, $\Lambda(\rho) \equiv \dfrac{\mathcal{L}(\rho)}{\max_{\rho'}\left[\mathcal{L}(\rho')\right]}$

4. Take the negative log, $\lambda(\rho) \equiv -2\log\Lambda(\rho)$



5. The confidence region is:

$$\boxed{\hat{\mathcal{R}}(D) = \left\{\text{all } \rho \text{ s.t. } \lambda(\rho) \leq \lambda_{\text{cutoff}(\alpha,k)}\right\}}$$

...where λ_{cutoff} is a function of the desired confidence (α), and $k = \min(d^2 - 1, \#\text{ of independent observables measured})$

The End

~~The End~~

I left out “Why” and “How”

~~The End~~

I left out "Why" and "How"

- Probability-ratio (PR) confidence regions optimize *average* power for a specified averaging measure.
- Seeking the *minimax* estimator (best worst-case behavior) yields *likelihood ratio* confidence regions.
- How to <u>describe</u> and <u>use</u> a confidence region.
- How to calculate the threshold, λ_{cutoff}.

Designing Optimal Confidence Region Estimators

Desiderata

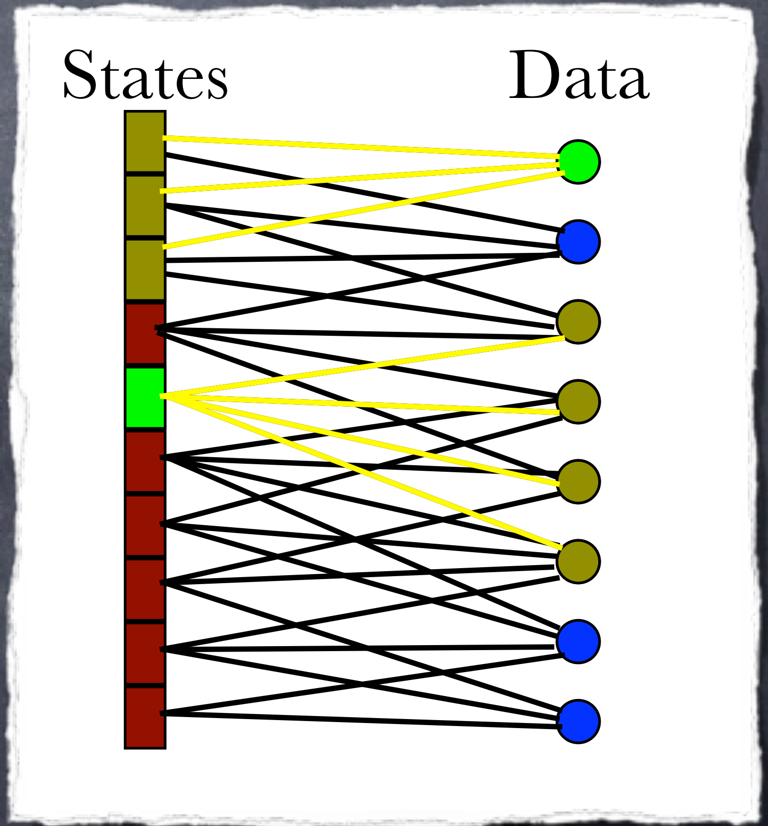
- CORRECTNESS: For every ρ, the total probability of datasets D for which $\hat{\mathcal{R}}(D)$ contains ρ must be at least α.
- POWER: $\hat{\mathcal{R}}(D)$ should be as small as possible.
- CONVENIENCE: $\hat{\mathcal{R}}(D)$ should be connected and (if possible) convex.

1. Correctness

- For every ρ, the total probability of datasets D so that $\hat{\mathcal{R}}(D)$ contains ρ must be at least α.

$$Pr(\mathcal{R}(D) \ni \rho|\rho) = \sum_{D:\ \mathcal{R}(D)\ni\rho} Pr(D|\rho) \geq \alpha$$

- Let ρ be connected to D iff $\rho \in \hat{\mathcal{R}}(D)$.
- Each ρ must be connected to datasets with total probability α.
- So ... which ones?
  - We choose the connections to minimize the *volume* of $\hat{\mathcal{R}}(D)$.

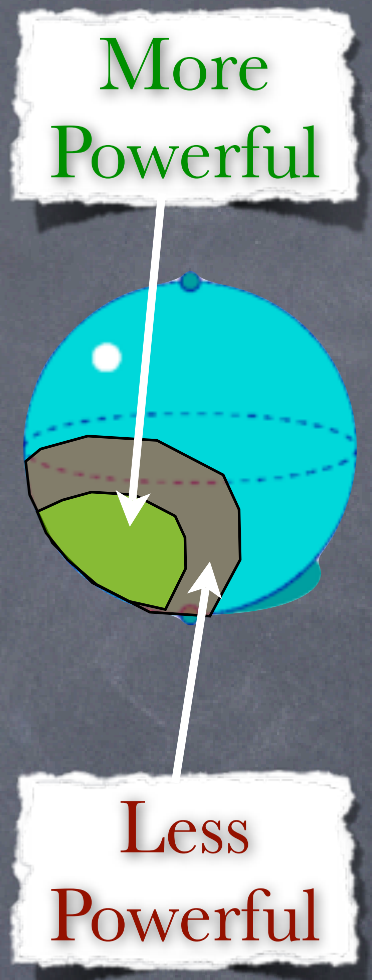


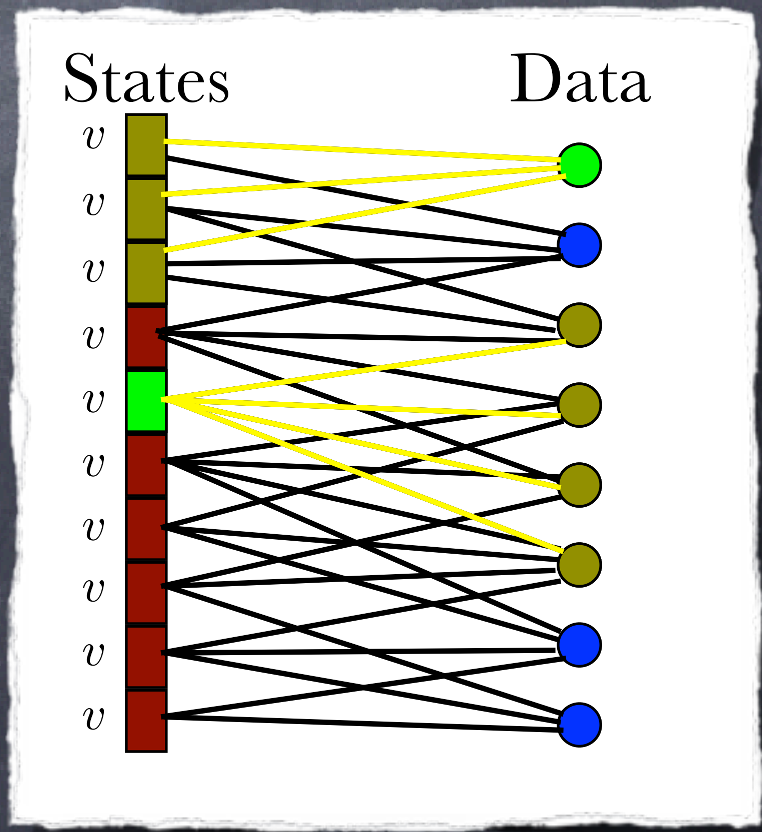
2. Power

- We'll choose connections
  to minimize "typical" volume of $\hat{\mathcal{R}}(D)$.

- PROBLEMS:
  1. No unique way to quantify volume.
  2. We can't simultaneously minimize
     volume for every dataset.
  3. In fact, we can't simultaneously minimize
     expected volume for more than one true state.

- But we *can* minimize the *average* expected
  volume of $\hat{\mathcal{R}}(D)$, w/r.t. any measure over ρ!



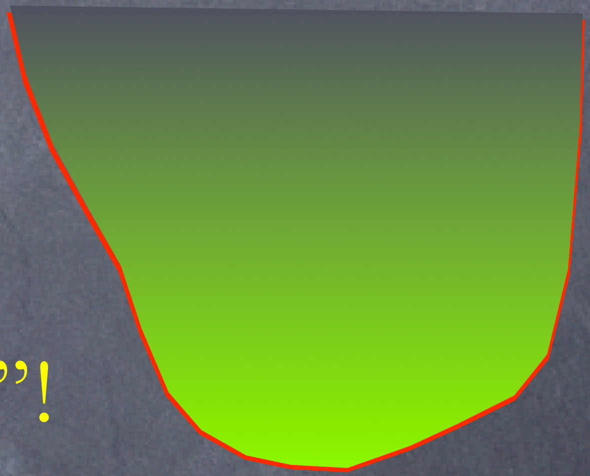


3. *Average* Power

- *Probability ratio* (PR) estimators (next slide) minimize the <u>average</u> of

$$\overline{V}(\rho) = \sum_D Pr(D|\rho) V(\hat{\mathcal{R}}(D))$$

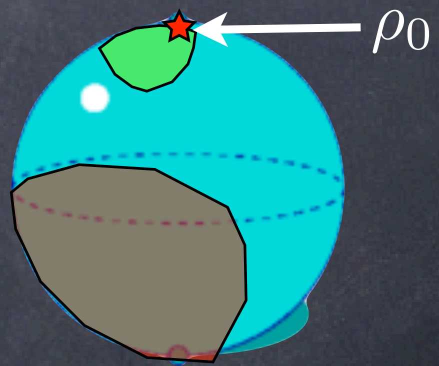
w/respect to a measure $Pr(\rho)\mathrm{d}\rho$.



- 1. Correctness does <u>not</u> depend on this "prior"!
  2. Every "best" estimator is of this form.
  3. Optimality holds for *every* volume measure!!!

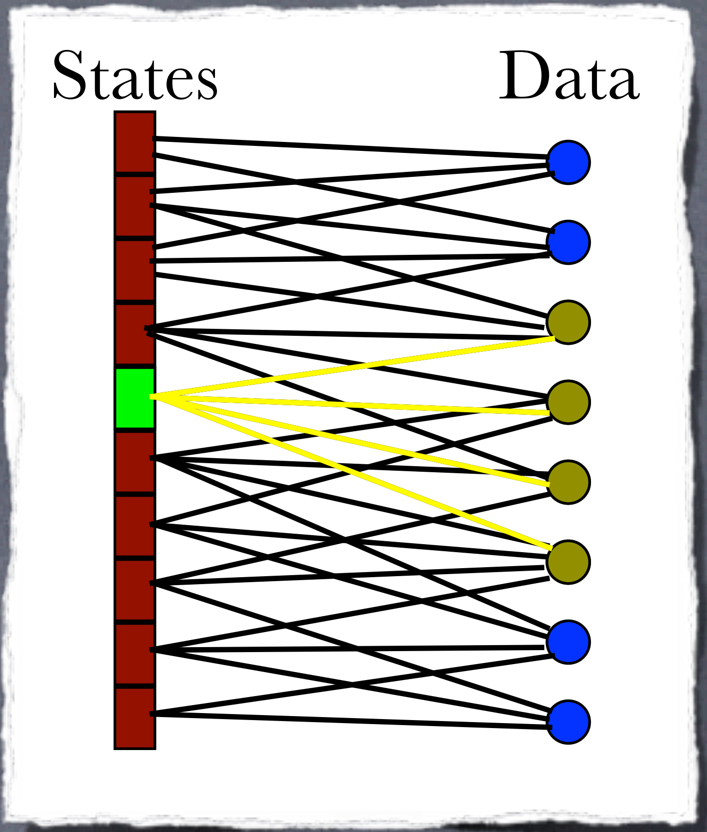
- **Example**: A brash experimentalist chooses $Pr(\rho)\mathrm{d}\rho = \delta(\rho - \rho_0)$

Effect: Somewhat smaller regions for typical D ***if*** $\rho = \rho_0$ ... much larger regions otherwise!



- **Result**: Consequences fall entirely on the experimentalist -- *NOT* on the end user of his conclusions!

4. *Probability Ratio* regions



- For each state ρ, we rank the datasets by an "attractiveness" function:

$$r_\rho(D) \equiv \frac{Pr(D|\rho)}{Pr(D)}$$

where $Pr(D) \equiv \int Pr(D|\rho)Pr(\rho)\mathrm{d}\rho$

- Each state gets connected to the *most* attractive states (highest *r*), until it's connected to datasets w/total probability α.

- Amazingly, this rule minimizes average expected volume -- according to *any* volume measure!!!

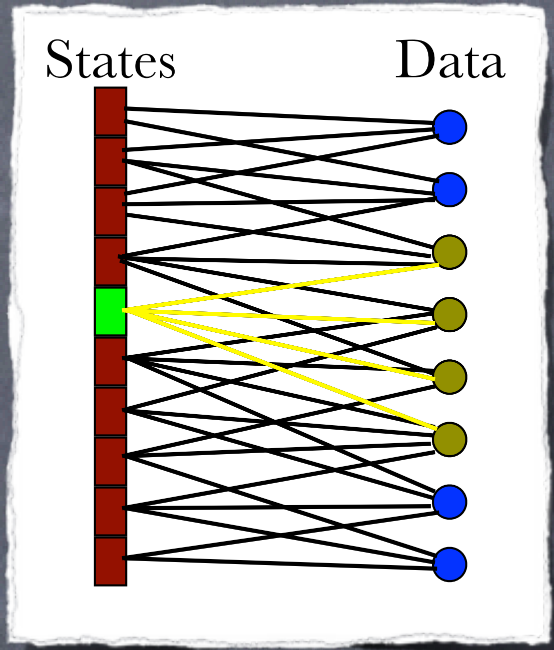
$$\langle\overline{V}\rangle = \int \overline{V}(\rho)Pr(\rho)\mathrm{d}\rho$$

5. *Worst Case* Power

- We connect ρ to the datasets with the highest

$$\frac{Pr(D|\rho)}{Pr(D)}$$

...but what $Pr(D)$ should we pick?



- GOAL: defend against <u>every</u> ρ (worst case).

- If $Pr(D')$ is low, then $\hat{\mathcal{R}}(D')$ will be large!...
  ...But that's bad if D' happens often.

- IMPLICATION: $Pr(D')$ should <u>only</u> be small if there is <u>no</u> ρ for which $Pr(D'|\rho)$ is large.

$$\implies Pr(D) \propto \max_{\rho} Pr(D|\rho) \implies \frac{Pr(D|\rho)}{Pr(D)} = \frac{\mathcal{L}(\rho)}{\mathcal{L}_{max}}$$

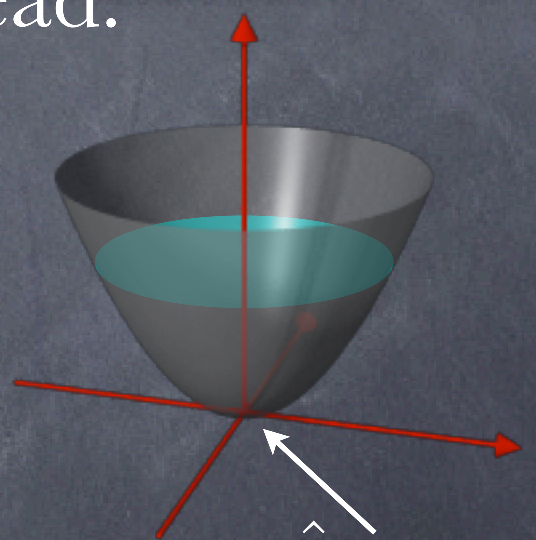
6. *Likelihood Ratio* regions

- Connect each ρ to datasets with high likelihood ratios:

$$\Lambda_D(\rho) \equiv \frac{\mathcal{L}(\rho)}{\mathcal{L}_{max}}$$

- It's convenient to use the loglikelihood-ratio instead:

$$\lambda_D(\rho) = -2\log\Lambda_D(\rho) \geq 0$$



So: $\rho \sim D$ iff $\lambda_D(\rho) < \lambda_{\text{cutoff}}(\rho)$, where $\lambda_{\text{cutoff}}(\rho)$ is defined by the correctness condition:

$$\sum_{D:\ \lambda_D(\rho)<\lambda_{\text{cutoff}}(\rho)} Pr(D|\rho) \gtrsim \alpha$$

- Finally, we invert the relationship to get $\hat{\mathcal{R}}(D)$:

$$\hat{\mathcal{R}}(D) = \{\text{all } \rho \text{ such that } \lambda_D(\rho) \leq \lambda_{\text{cutoff}}(\rho)\}$$

- Remaining technical problem: determine $\lambda_{\text{cutoff}}(\rho)$.

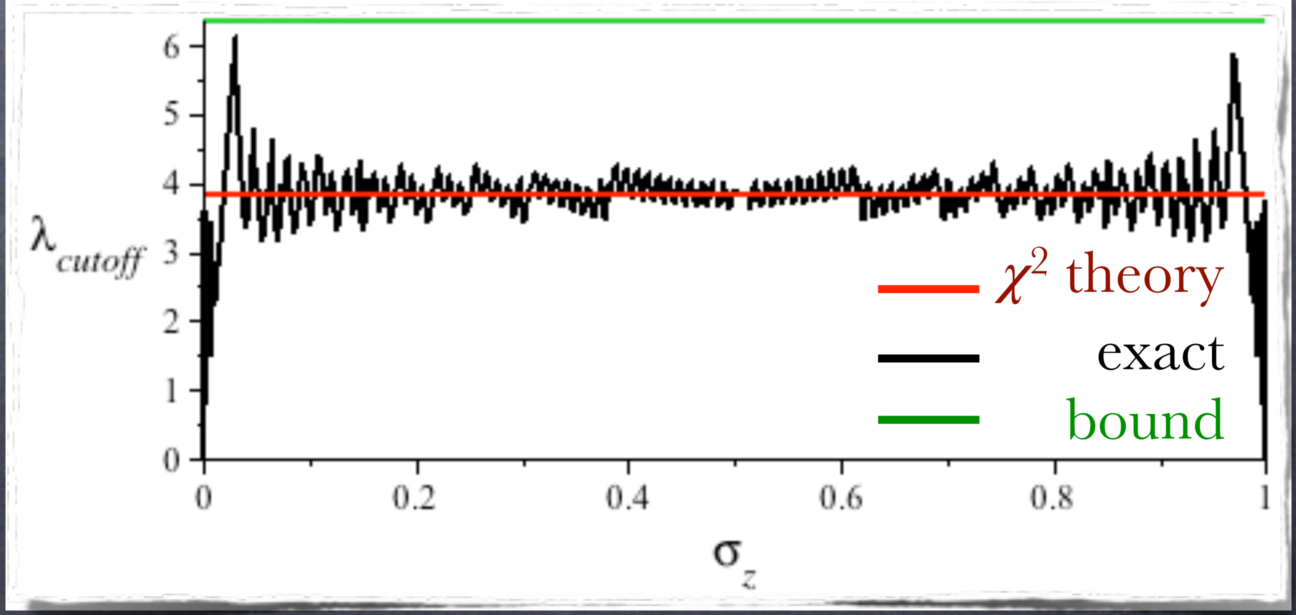
The Cutoff

- Cutoff $\Rightarrow$ size of regions $\Rightarrow$ correctness & power:
  Too low? NOT CORRECT   Too high? NOT POWERFUL
- Set $\lambda_{\text{cutoff}}(\rho)$ based on the complementary cumulative distribution function of λ, given ρ:

$$F(x) \equiv 1 - \int_0^x Pr(\lambda)\mathrm{d}\lambda$$

$$F(\lambda_{\text{cutoff}}) = 1 - \alpha$$

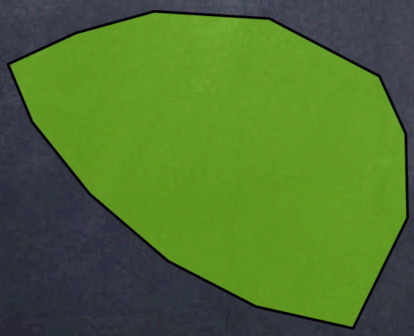
- The state-dependent cutoff fluctuates...
  $\Rightarrow$ $\hat{\mathcal{R}}(D)$ not convex!



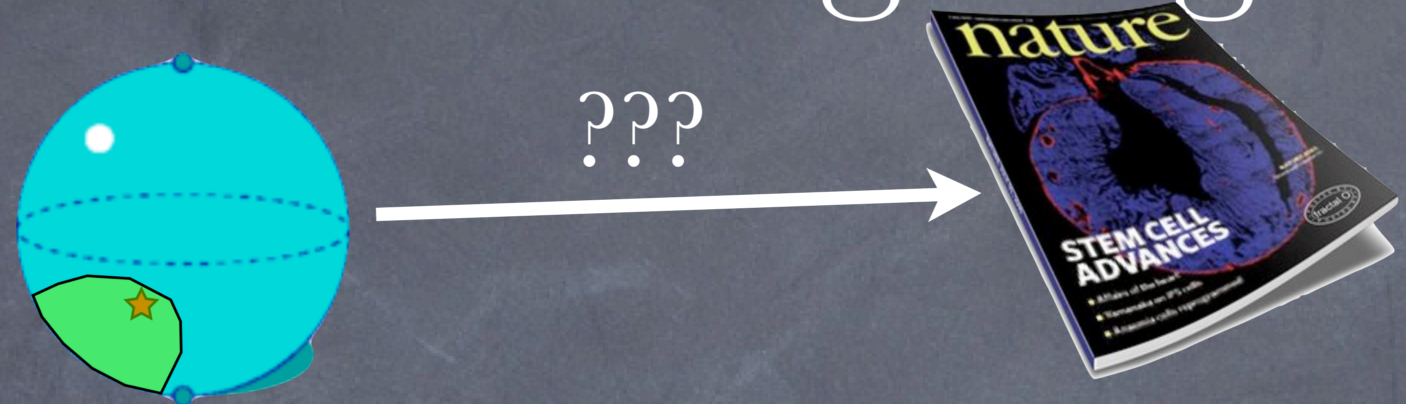
- So we use a constant upper bound λ_{cutoff} instead.

Describing Regions

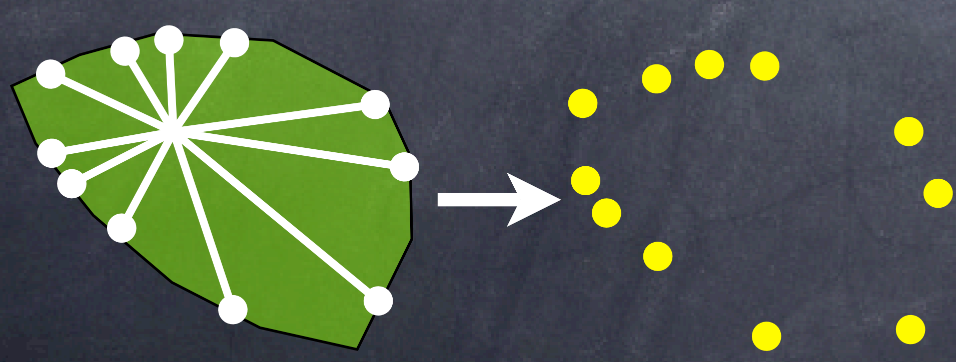
- Likelihood region confidence regions are about as convenient as possible...
- ...But how do you describe an arbitrary convex region in density matrix space??



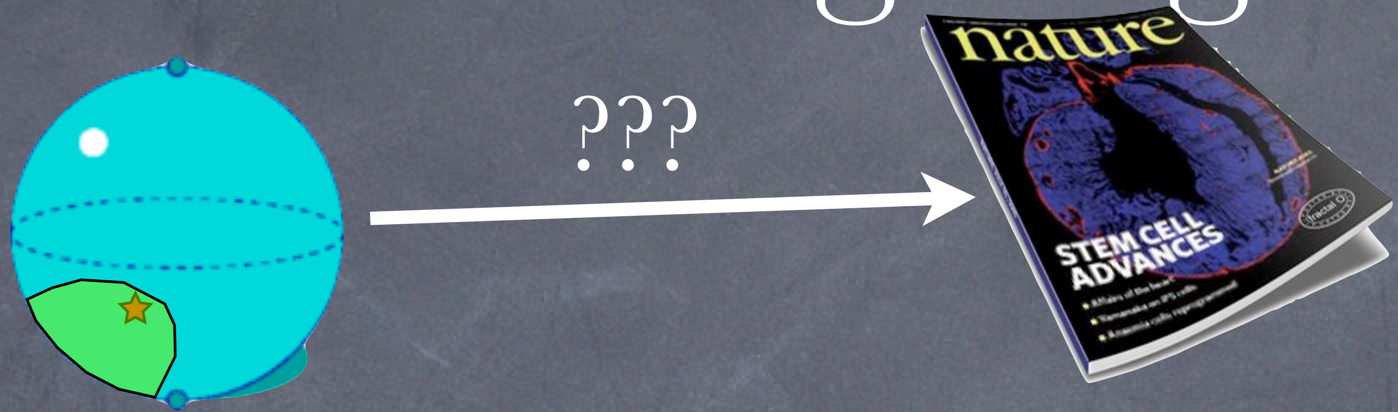
Describing Regions



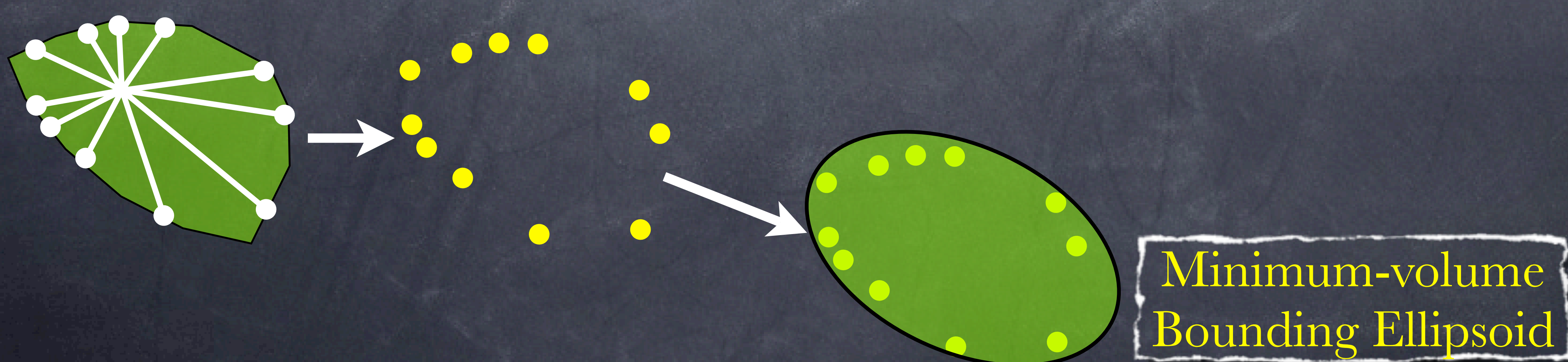
- Likelihood region confidence regions are about as convenient as possible...
- ...But how do you describe an arbitrary convex region in density matrix space??



Describing Regions

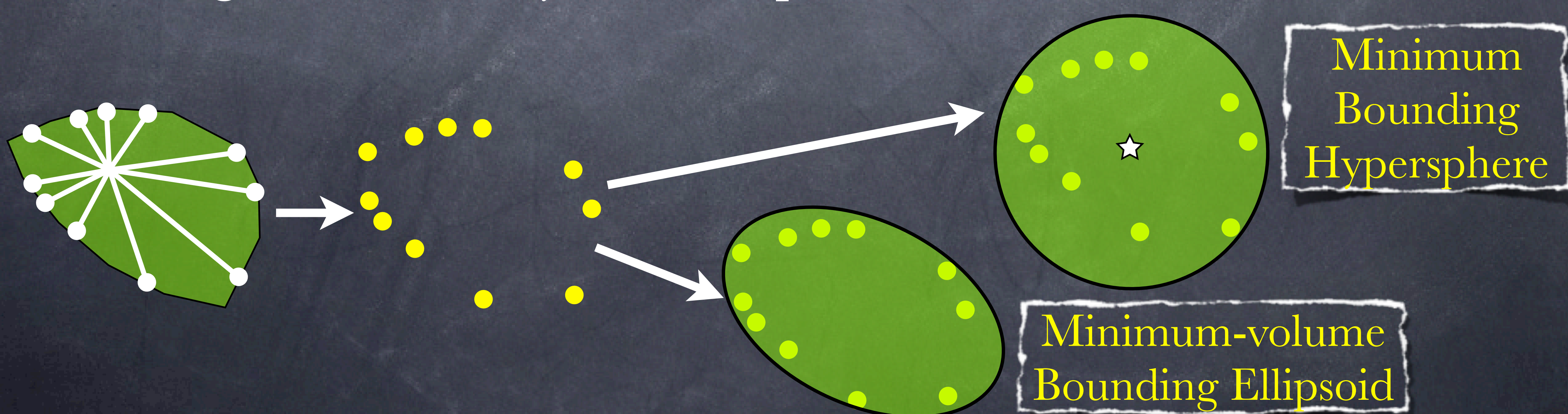


- Likelihood region confidence regions are about as convenient as possible...
- ...But how do you describe an arbitrary convex region in density matrix space??



Describing Regions

- Likelihood region confidence regions are about as convenient as possible...
- ...But how do you describe an arbitrary convex region in density matrix space??



CRs in daily lab/theory life

- Why an *experimentalist* should estimate CRs every day.
  - You want to estimate the error in tomography... ...but you don't know the true state!
  - You want to figure out which tomographic procedure is the most reliable.
- Why a *theorist* should think about CRs every day.
  - The LR construction points to a *robust* analytic generalization of Fisher information -- which faithfully describes uncertainty about states.

Conclusions

- Accurate region estimators are:
  1. What error bars want to be when they grow up.
  2. Better than point estimates (e.g. for fault tolerance)
  3. A surprisingly useful conceptual tool, too!
- Likelihood ratio confidence regions are:
  1. Conceptually simple and elegant
  2. Absolutely reliable <u>and</u> near-optimal in power.
- Cutoff is understood... but we need better bounds.
- Describing a confidence region is nontrivial...
  ...but actually fairly tractable.

Setting the Cutoff

The Role of the Cutoff

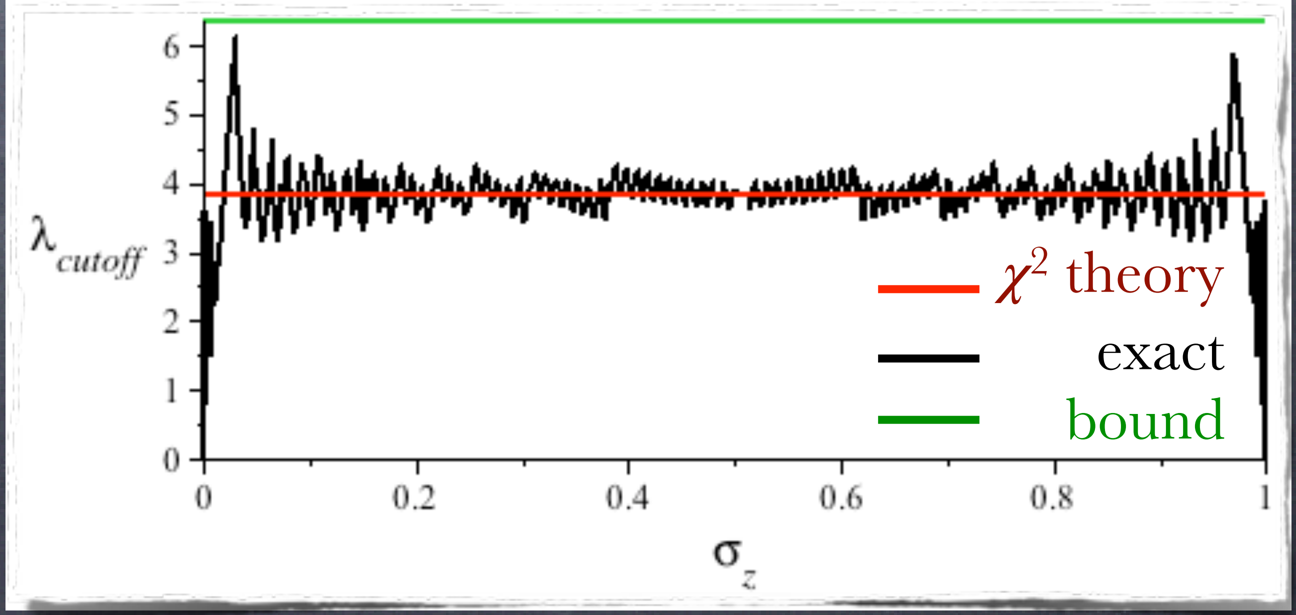
- Cutoff $\Rightarrow$ size of regions $\Rightarrow$ correctness & power:
  Too low? NOT CORRECT Too high? NOT POWERFUL

- Set $\lambda_{\text{cutoff}}(\rho)$ based on the complementary cumulative distribution function of λ, given ρ:

$$F(x) \equiv 1 - \int_0^x Pr(\lambda) \mathrm{d}\lambda$$

$$F(\lambda_{\text{cutoff}}) = 1 - \alpha$$

- The state-dependent cutoff fluctuates...
  $\Rightarrow \hat{\mathcal{R}}(D)$ not convex!

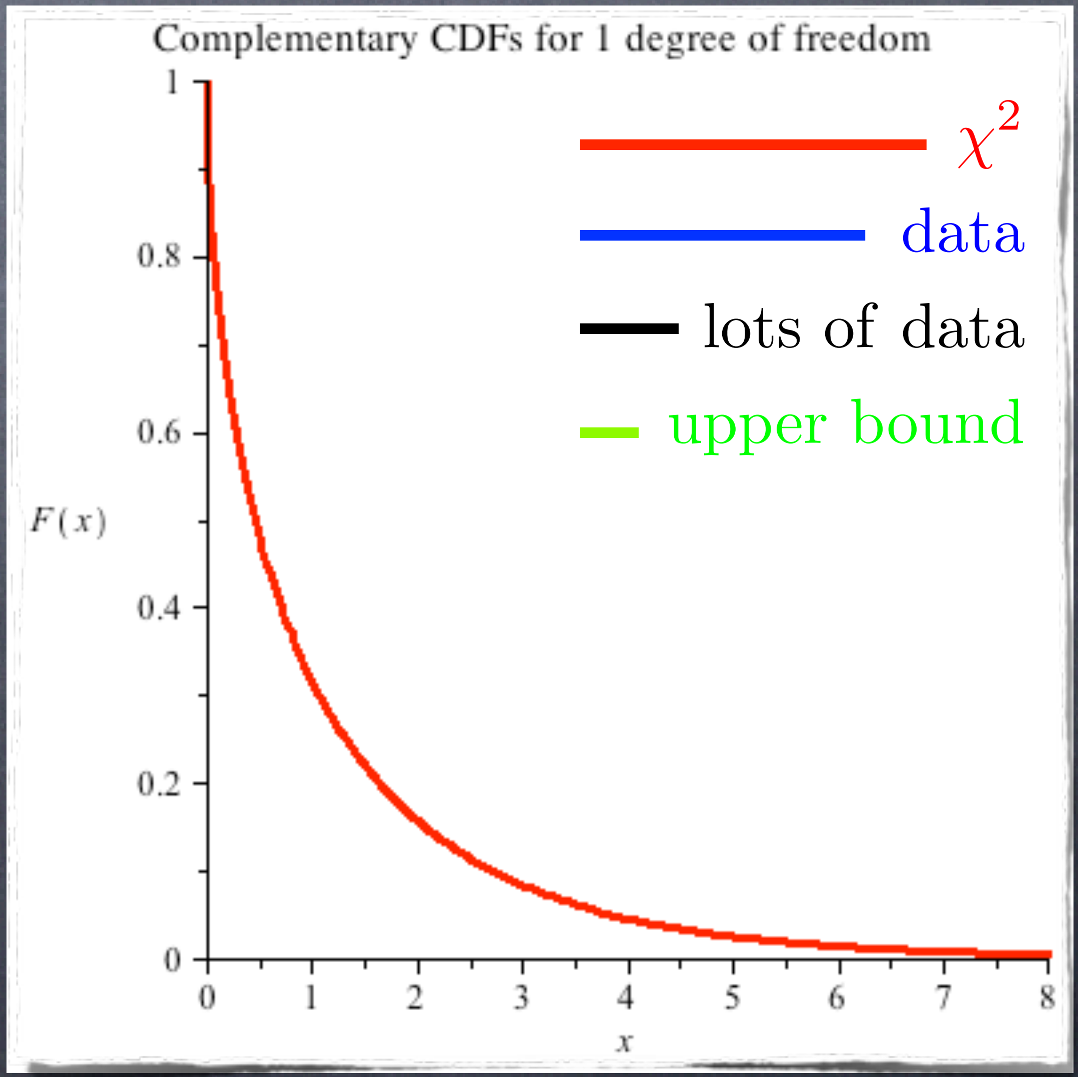


- To avoid this, we can use a constant upper bound λ_{cutoff} instead.

χ^2 theory does not work

- If the data were Gaussian, with K degrees of freedom, then λ would be a χ^2_K variable:

$$
\begin{aligned}
F_{\chi^2_1}(x) &= \operatorname{erfc}\left(\sqrt{x/2}\right) \\
F_{\chi^2_2}(x) &= e^{-x/2} \\
&\ldots \\
F_{\chi^2_K}(x) &\approx \sqrt{x}^{K-2} e^{-x/2}
\end{aligned}
$$



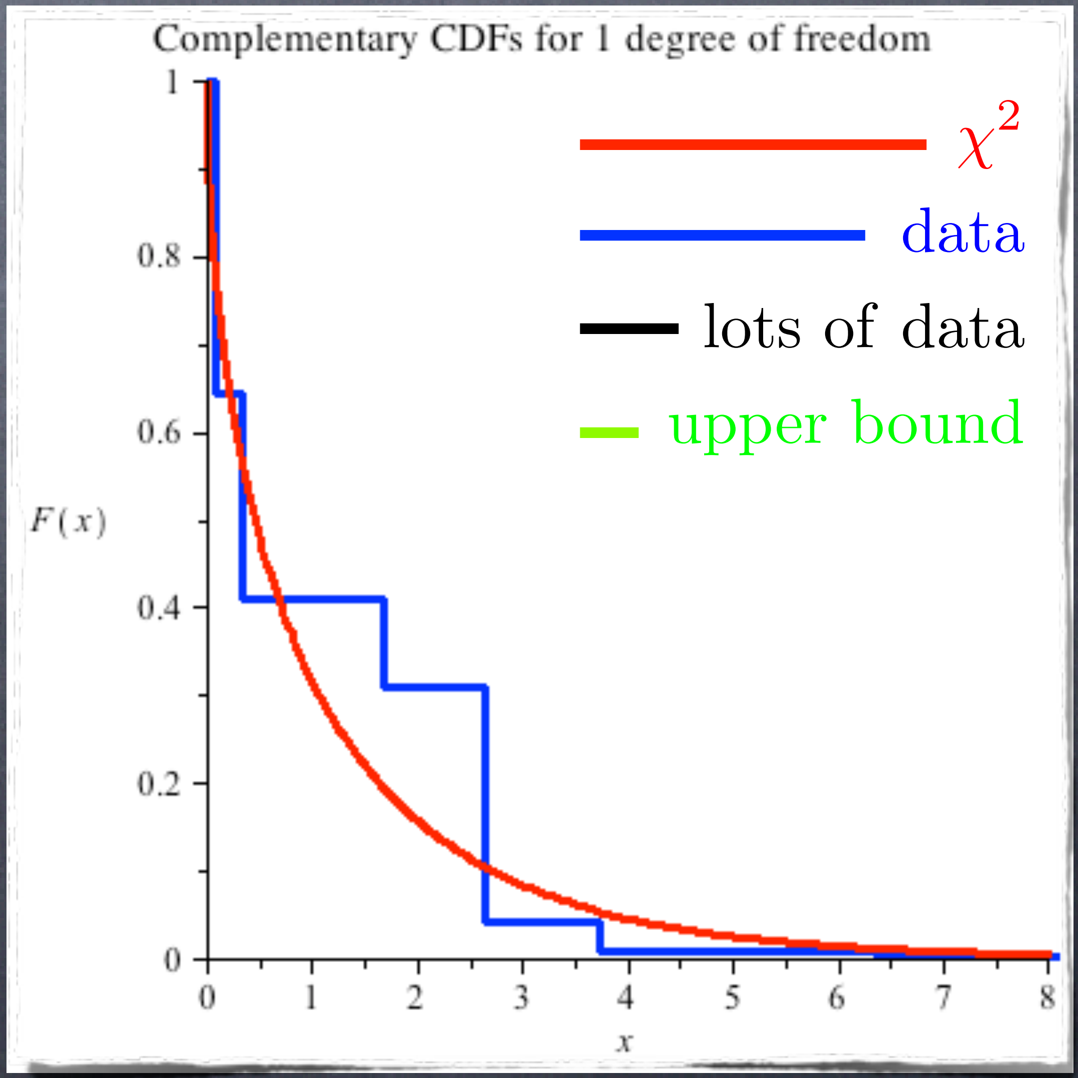
- Tomographic data is *not* Gaussian -- it is discrete.
- We need a valid upper bound for $F(x)$.

χ^2 theory does not work

- If the data were Gaussian, with K degrees of freedom, then λ would be a χ^2_K variable:

$$\begin{aligned} F_{\chi^2_1}(x) &= \operatorname{erfc}\left(\sqrt{x/2}\right) \\ F_{\chi^2_2}(x) &= e^{-x/2} \\ &\dots \\ F_{\chi^2_K}(x) &\approx \sqrt{x}^{K-2} e^{-x/2} \end{aligned}$$

- Tomographic data is *not* Gaussian -- it is discrete.
- We need a valid upper bound for $F(x)$.

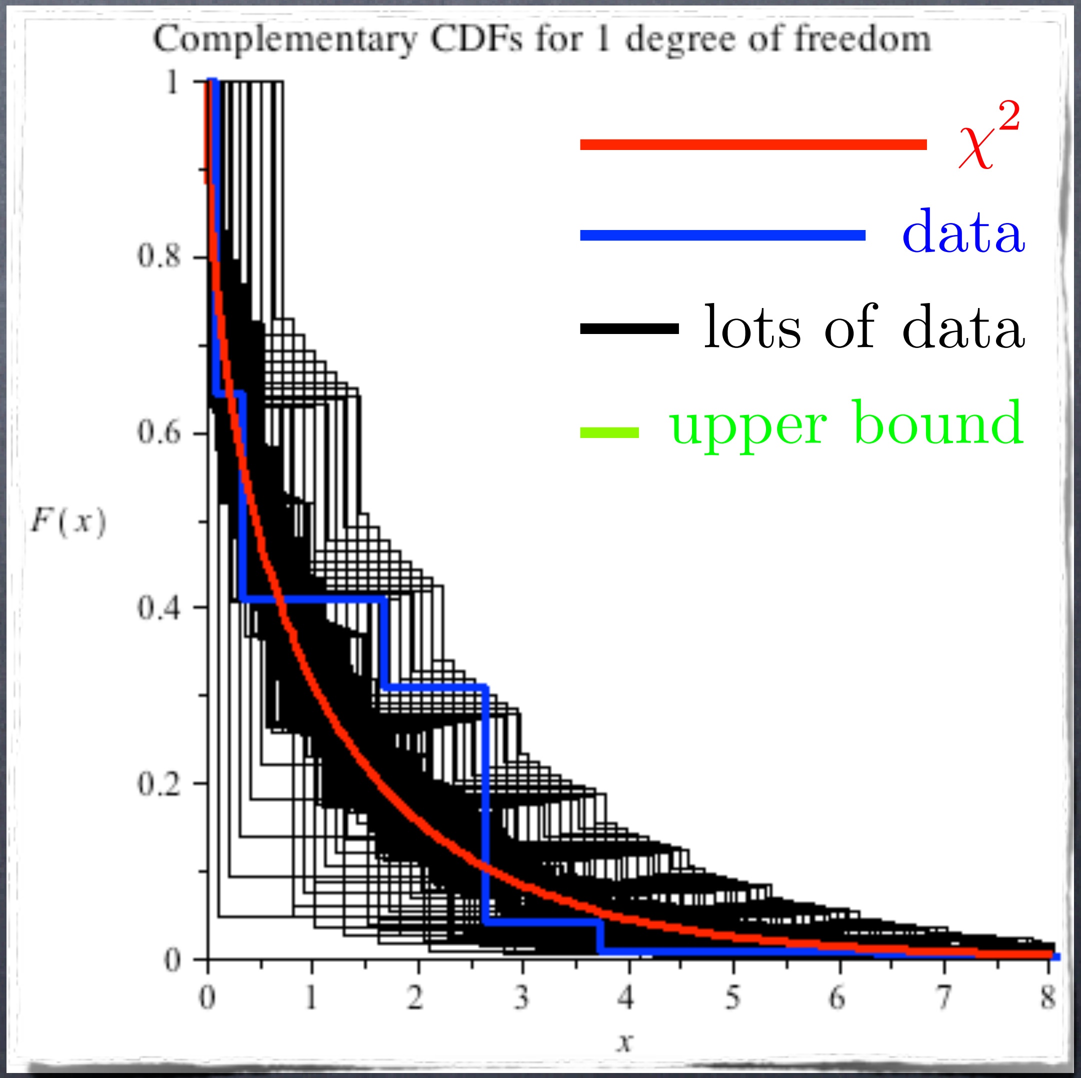


χ^2 theory does not work

- If the data were Gaussian, with K degrees of freedom, then λ would be a χ^2_K variable:

$$
\begin{aligned}
F_{\chi^2_1}(x) &= \operatorname{erfc}\left(\sqrt{x/2}\right) \\
F_{\chi^2_2}(x) &= e^{-x/2} \\
&\dots \\
F_{\chi^2_K}(x) &\approx \sqrt{x}^{K-2} e^{-x/2}
\end{aligned}
$$

- Tomographic data is *not* Gaussian -- it is discrete.
- We need a valid upper bound for $F(x)$.

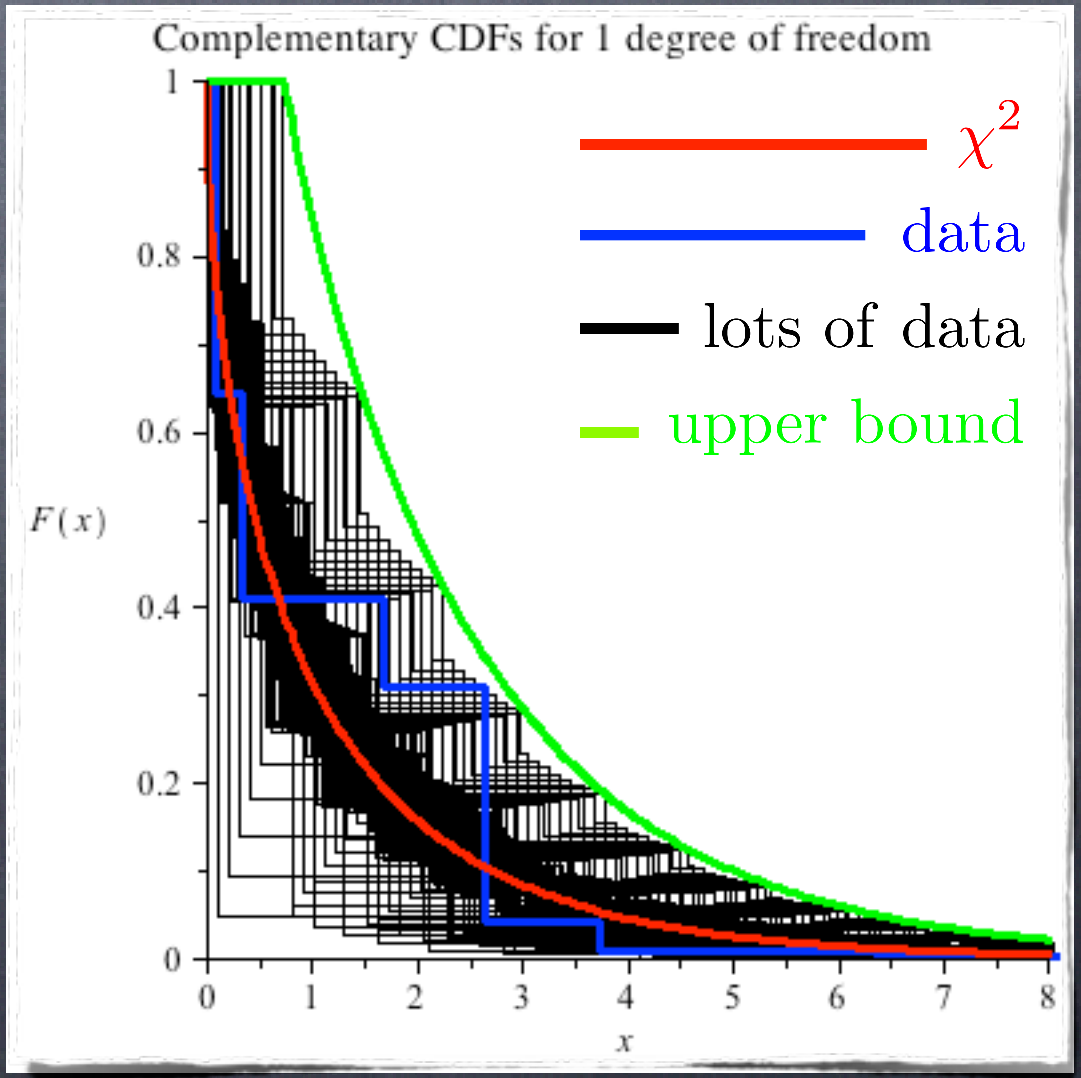


χ^2 theory does <u>not</u> work

- If the data were <u>Gaussian</u>, with K degrees of freedom, then λ would be a χ^2_K variable:

$$
\begin{aligned}
F_{\chi^2_1}(x) &= \operatorname{erfc}\left(\sqrt{x/2}\right) \\
F_{\chi^2_2}(x) &= e^{-x/2} \\
&\ldots \\
F_{\chi^2_K}(x) &\approx \sqrt{x}^{K-2} e^{-x/2}
\end{aligned}
$$

- Tomographic data is *not* Gaussian -- it is discrete.
- We need a valid upper bound for $F(x)$.



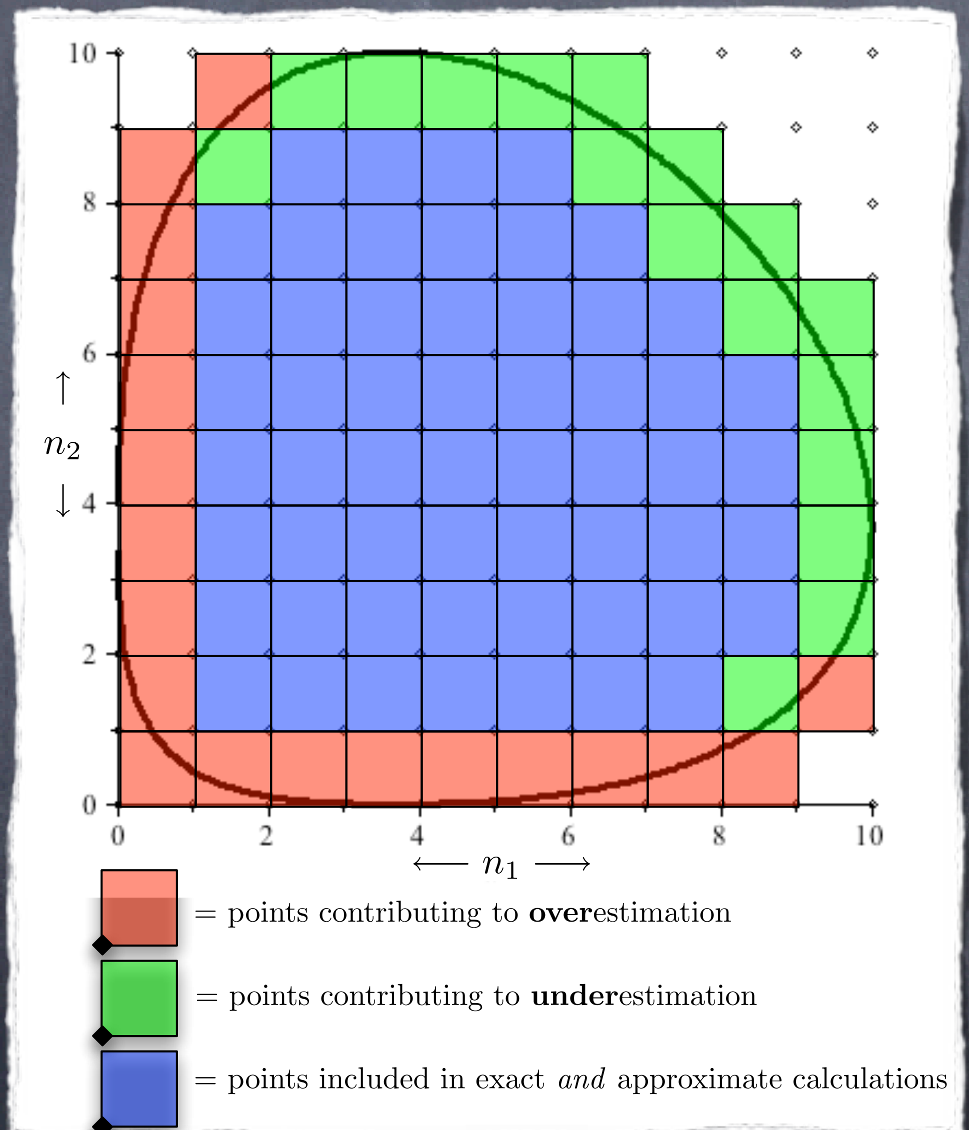
Deriving $F(x)$: a sketch

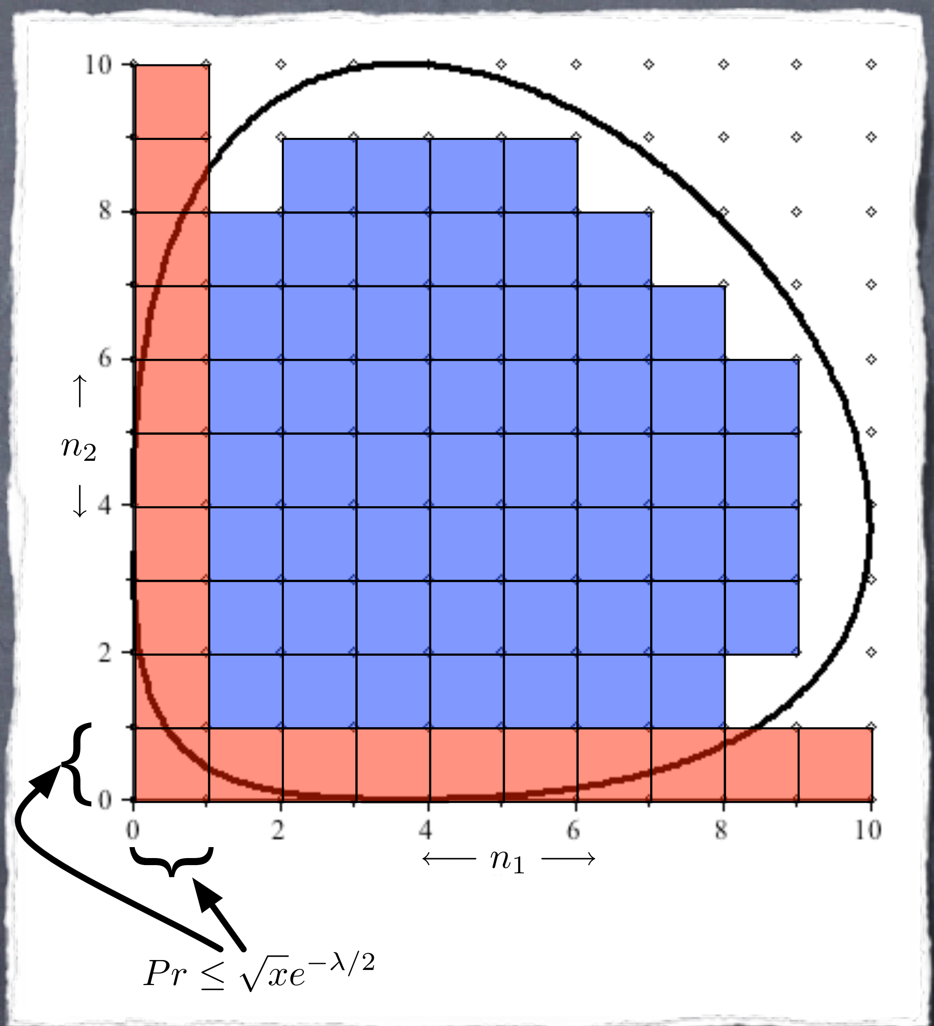
- Step 1: Tomographic data is essentially *multinomial*.
- Step 2: Approximate by multivariate *Poisson*.
- Step 3: Construct a "continuous Poisson" distribution for $\tilde{n}$, so that n=*floor($\tilde{n}$)* is Poisson.
- Step 4: Prove that for $\tilde{n}$, λ <u>is</u> (at worst) a χ^2_K variable.
- Step 5: So the discrepancy comes from *discreteness*.
- Step 6: Derive an upper bound on the extra $F(x)$ that results because tomographic data are discrete.

$$F_\lambda(x) = F_{\chi^2_K}(x) + F_{\mathrm{extra}}(x)$$

Correcting for discretion

- The extra probability comes from cells that lie [mostly] inside the $\lambda = x$ contour, but whose data point lies outside.



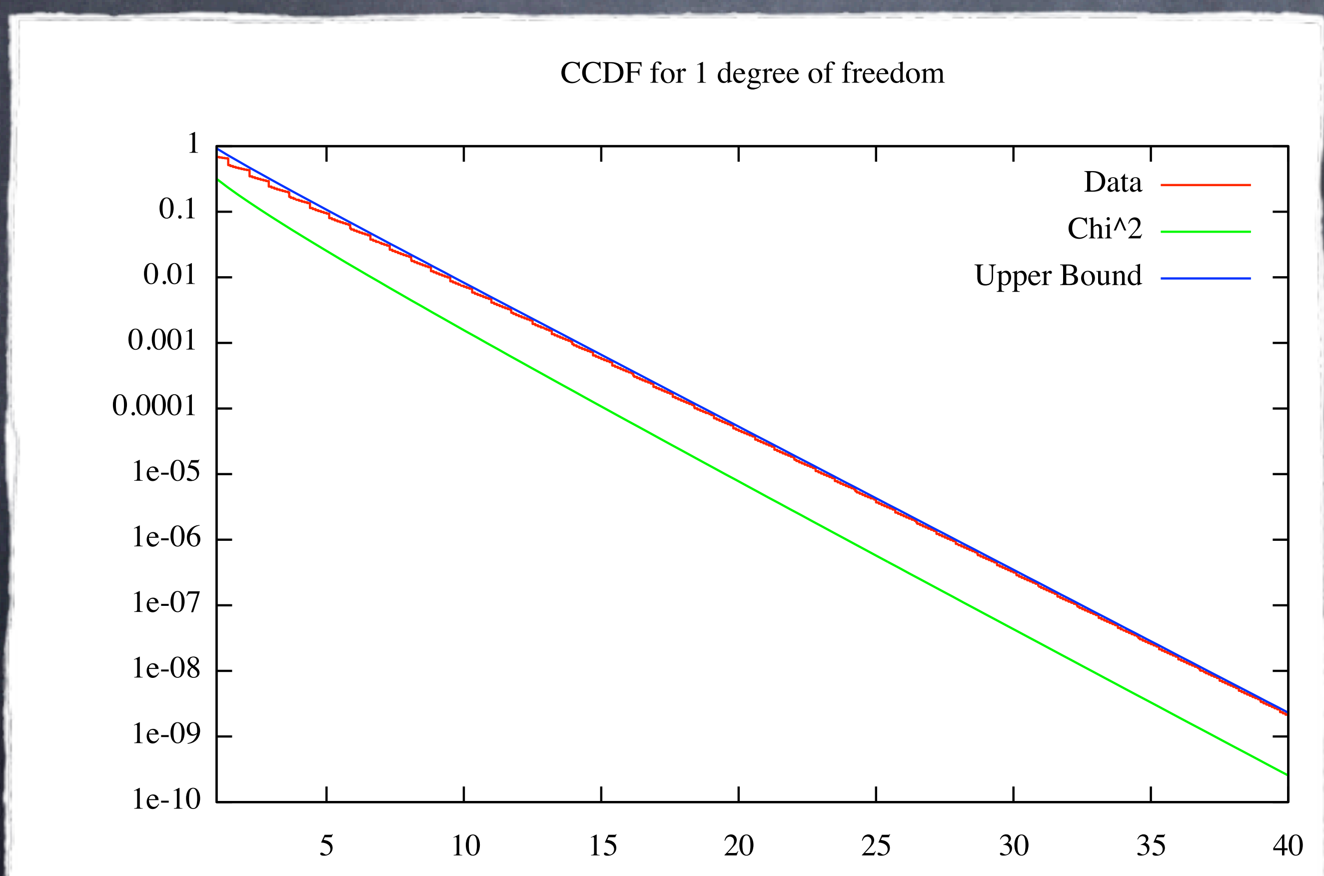


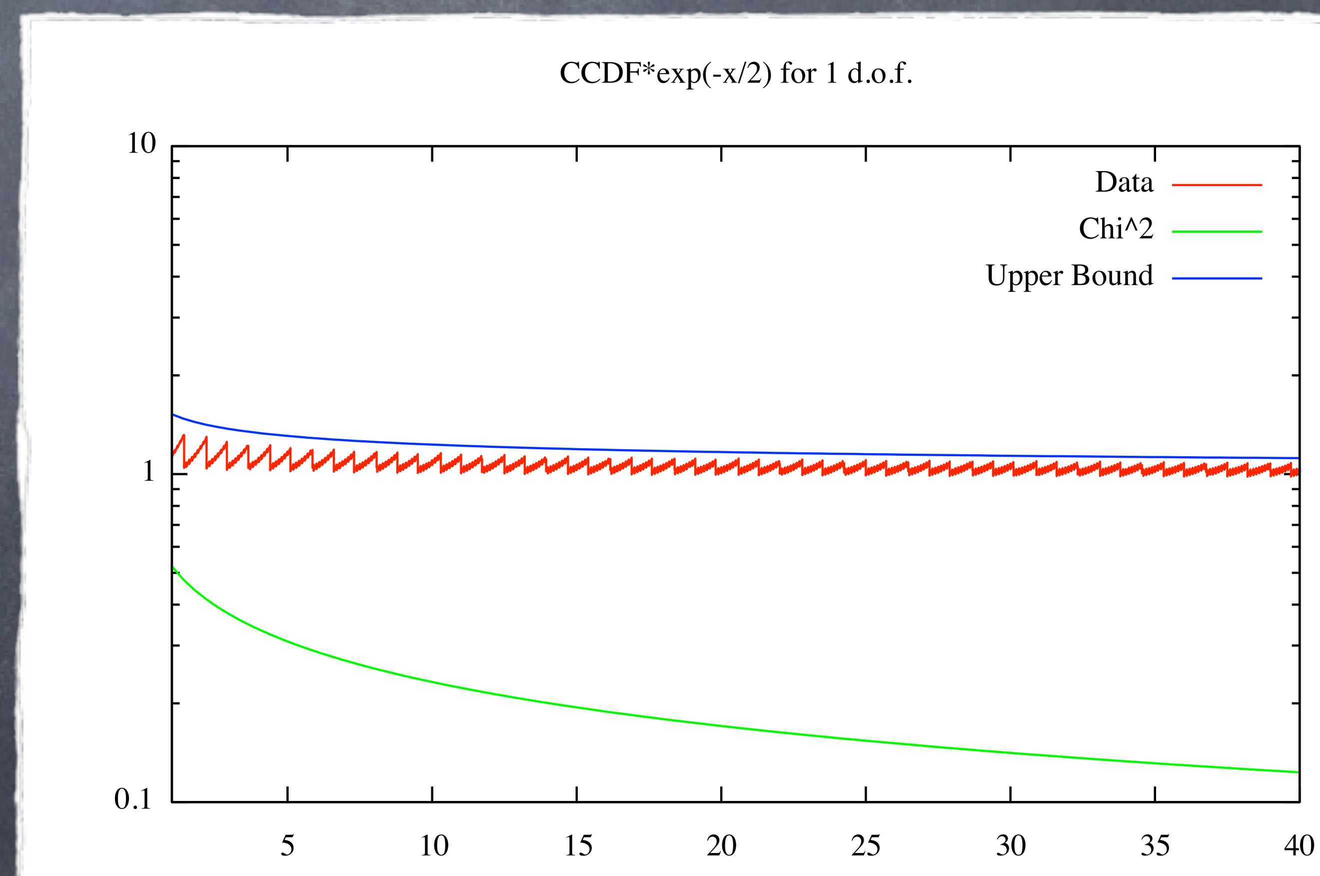
- Now simplify the boundary by pushing it out to the axes.

$$F_{\text{extra}} = Pr(\text{boundary}) \leq Pr(\text{sheath}) = \sum_{\text{sheath}} e^{-\frac{\lambda}{2}} Pr(\vec{n}|\vec{n})$$

A workable upper bound

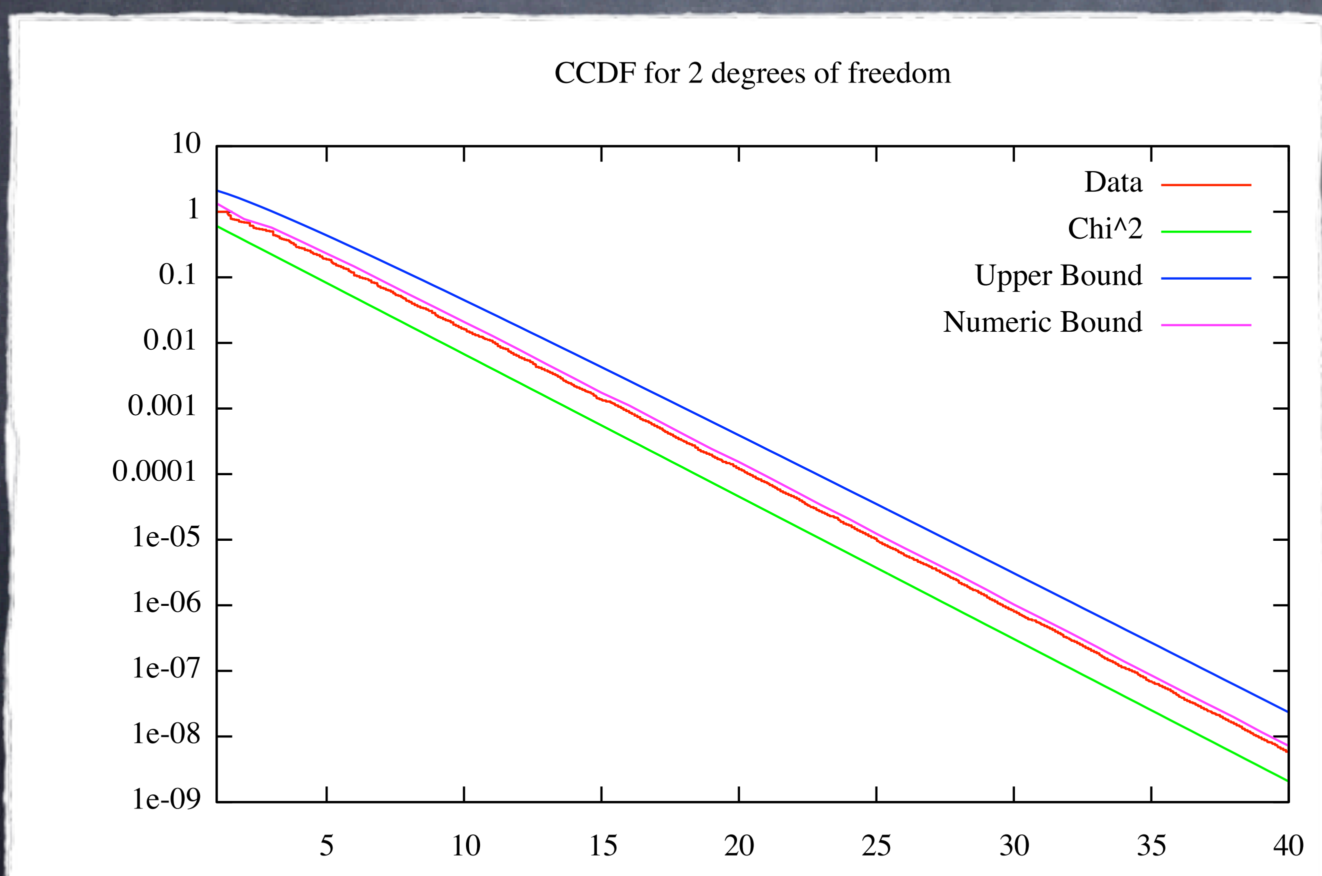
$$
\begin{aligned}
F(x) &\le F_{\chi^2_K} + e^{-x/2}\left[\left(1+\frac{\sqrt{3ex}}{\pi}\right)^K - \frac{\sqrt{3ex}}{\pi}^K\right] \\
&\approx F_{\chi^2_K} + e^{-x/2} K \sqrt{x}^{K-1},
\end{aligned}
$$

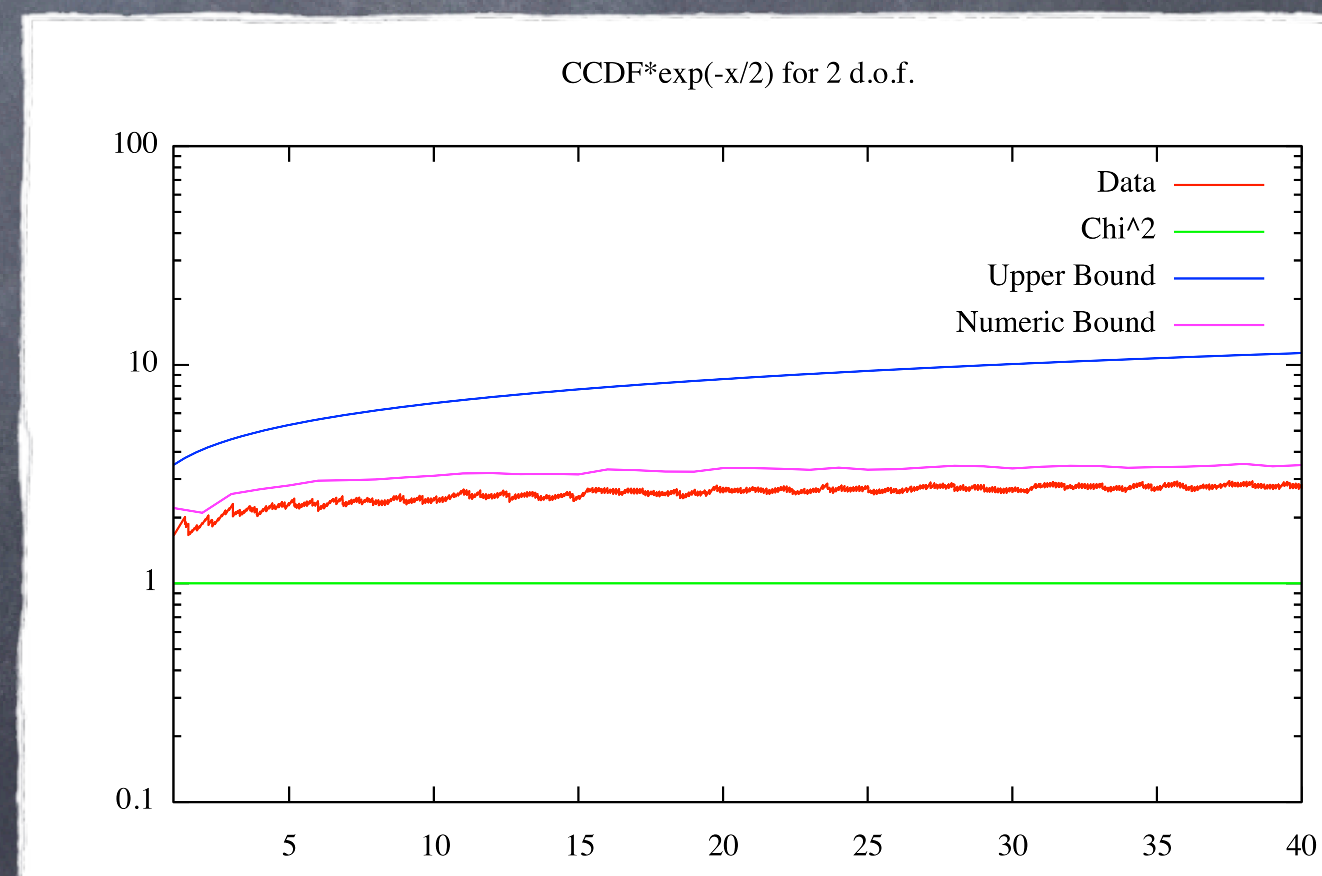




A workable upper bound

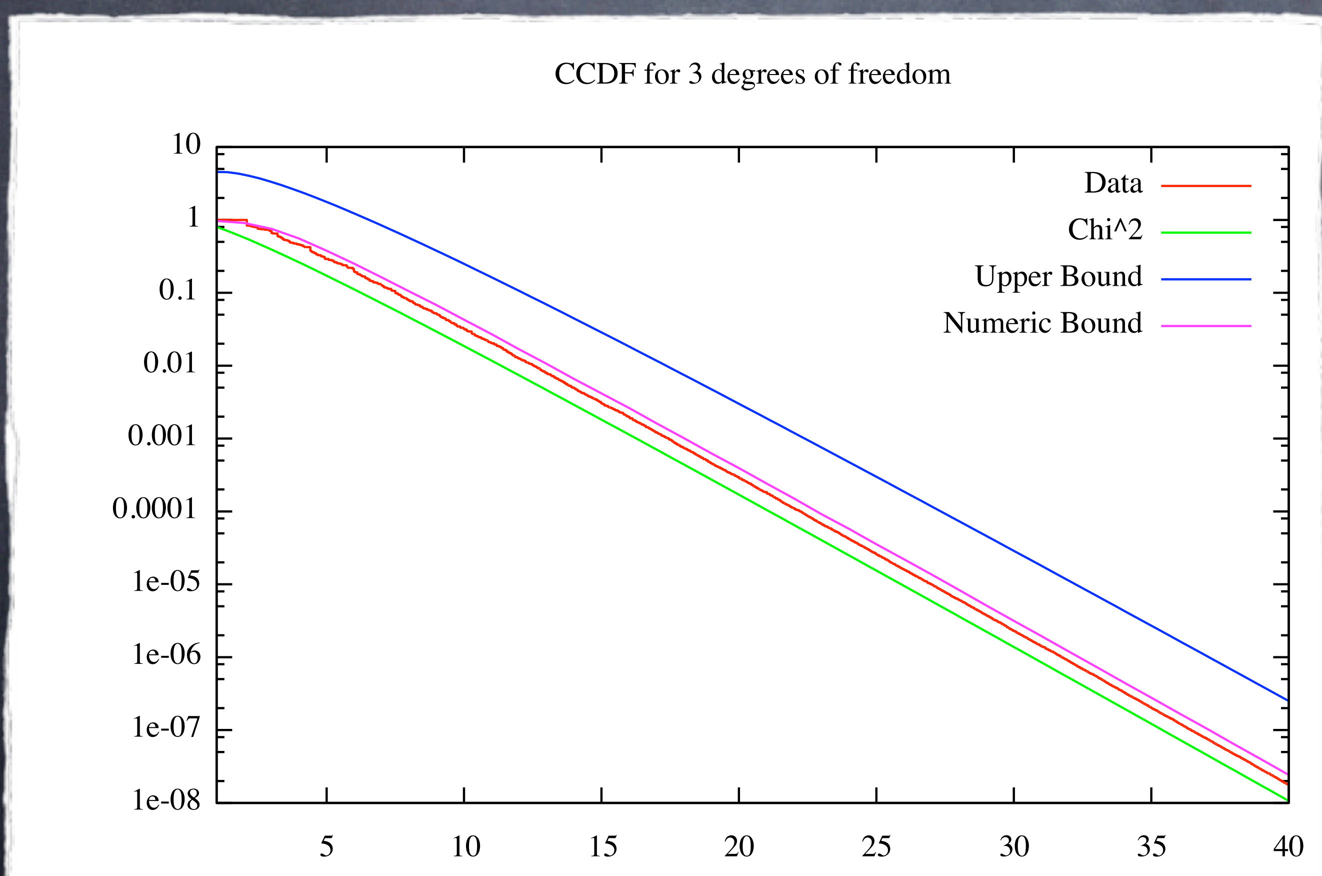
$$
\begin{aligned}
F(x) &\leq F_{\chi^2_K} + e^{-x/2}\left[\left(1+\frac{\sqrt{3ex}}{\pi}\right)^K - \frac{\sqrt{3ex}}{\pi}^K\right] \\
&\approx F_{\chi^2_K} + e^{-x/2} K \sqrt{x}^{K-1},
\end{aligned}
$$

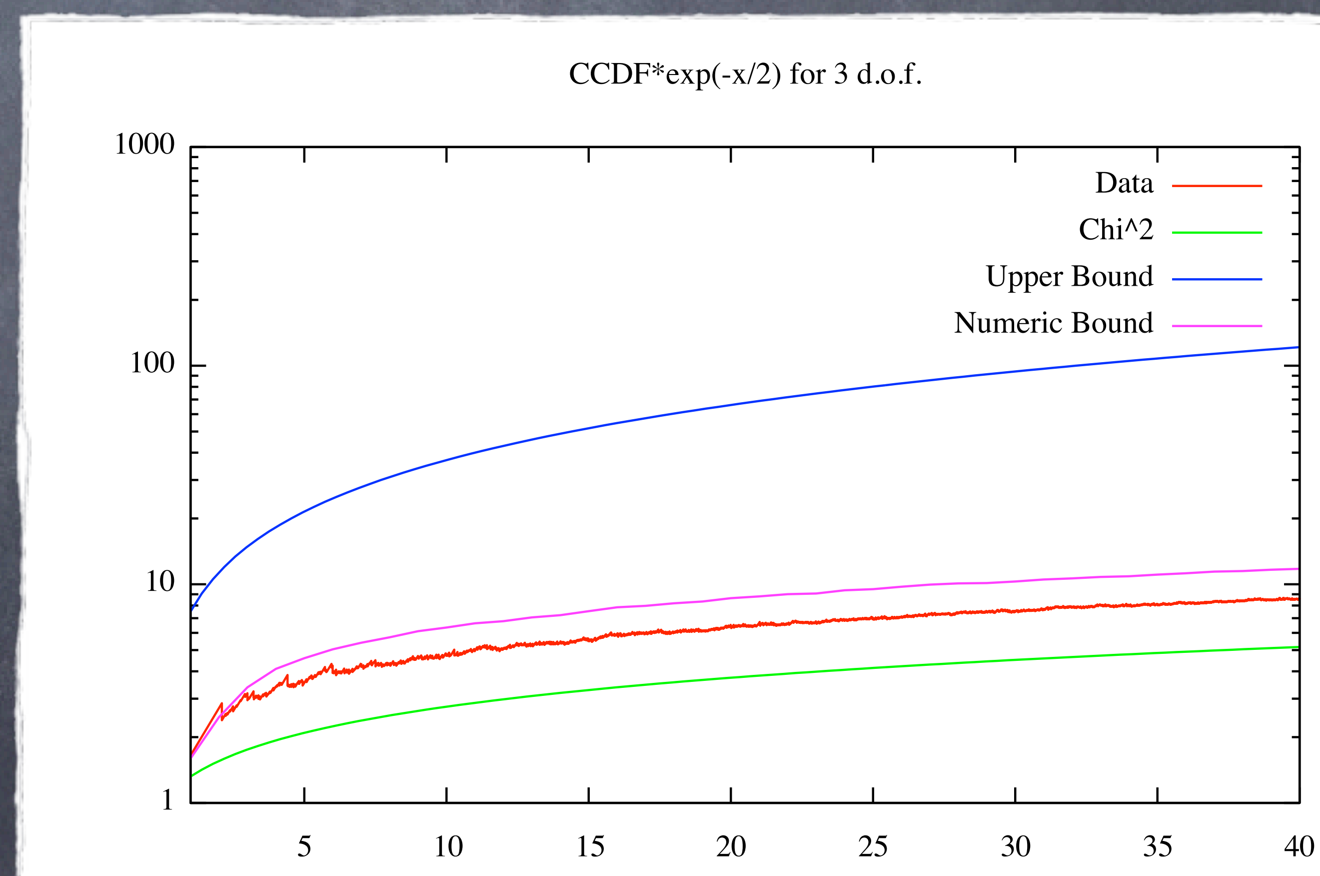




A workable upper bound

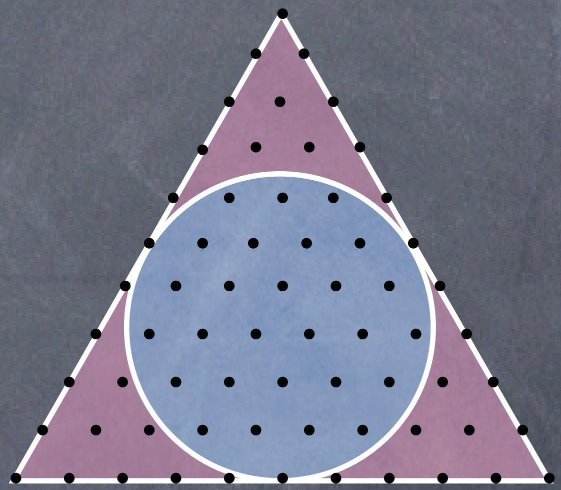
$$
\begin{aligned}
F(x) &\leq F_{\chi^2_K} + e^{-x/2}\left[\left(1+\frac{\sqrt{3ex}}{\pi}\right)^K - \frac{\sqrt{3ex}}{\pi}^K\right] \\
&\approx F_{\chi^2_K} + e^{-x/2} K \sqrt{x}^{K-1},
\end{aligned}
$$





Quantum Constraints

- Tomographic data is multinomially distributed...
- ...but positivity rules out many probability distributions -- especially near the boundary.



- This reduces λ for near-boundary states.
- For nearly-pure states, even χ^2 approximation is too conservative: nearly-pure states show up in $\hat{\mathcal{R}}(D)$ too often!