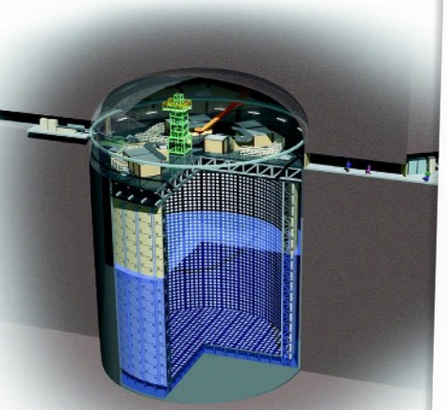
Efficient statistics for future neutrino oscillation experiments

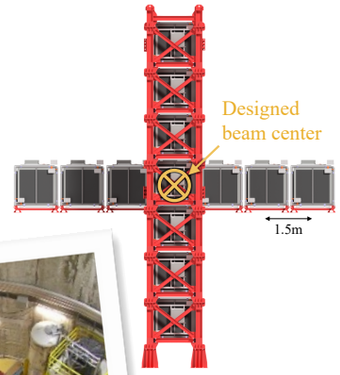
2022-06-19, Erice School of Subnuclear Physics
Lukas Berns, Tohoku University

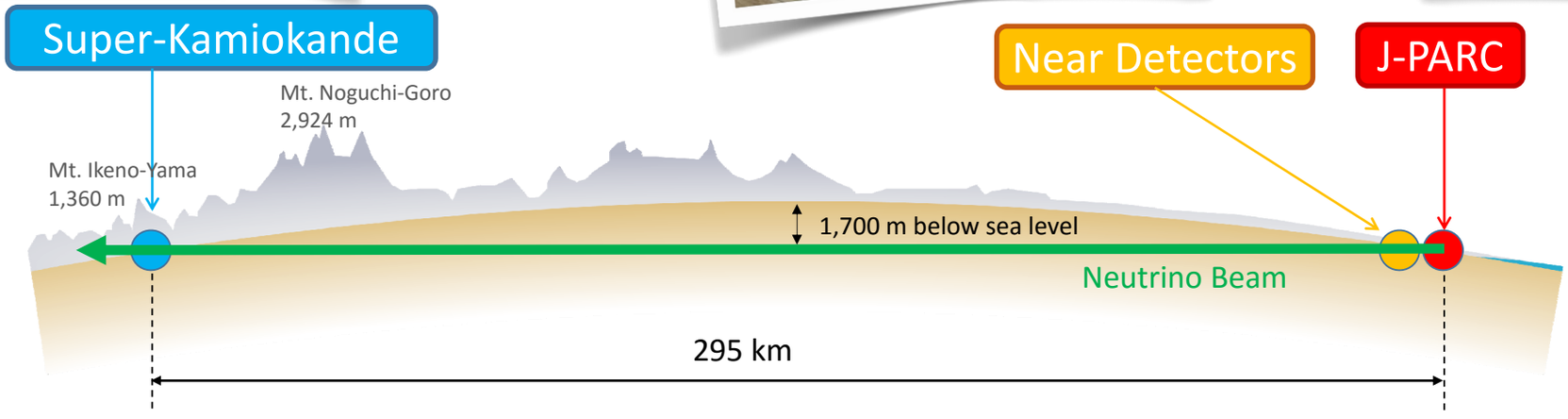
first few slides are with T2K and SK Collaborations

↑ view on Sendai city

T2K experiment







$\nu_\tau, \nu_\tau, \nu_\tau, \nu_\tau, \nu_e, \nu_\mu, \nu_\mu \leftarrow \nu_\mu, \nu_\mu, \nu_\mu, \nu_\mu, \nu_\mu, \nu_\mu, \nu_\mu$

- Study oscillation of neutrino beam from J-PARC accelerator
- ~500 collaborators from institutions in 12 countries

ν-oscillation

(interaction) (propagation)

For neutrinos flavor basis ≠ Hamiltonian basis.

→ Flavor (ν_e | ν_μ | ν_τ) oscillates over $L \times \Delta m^2 / E$,
amplitude controlled by (PMNS) **mixing matrix** U:

atmospheric — reactor — solar

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{CP}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{CP}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

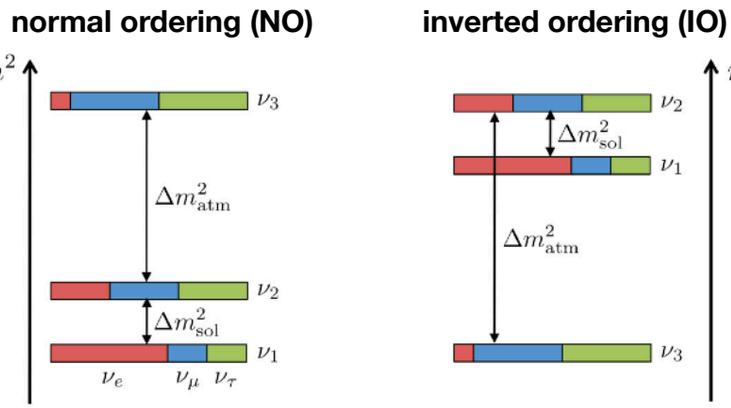
$c_{ij} \equiv \cos\theta_{ij}$
$s_{ij} \equiv \sin\theta_{ij}$

Open questions:

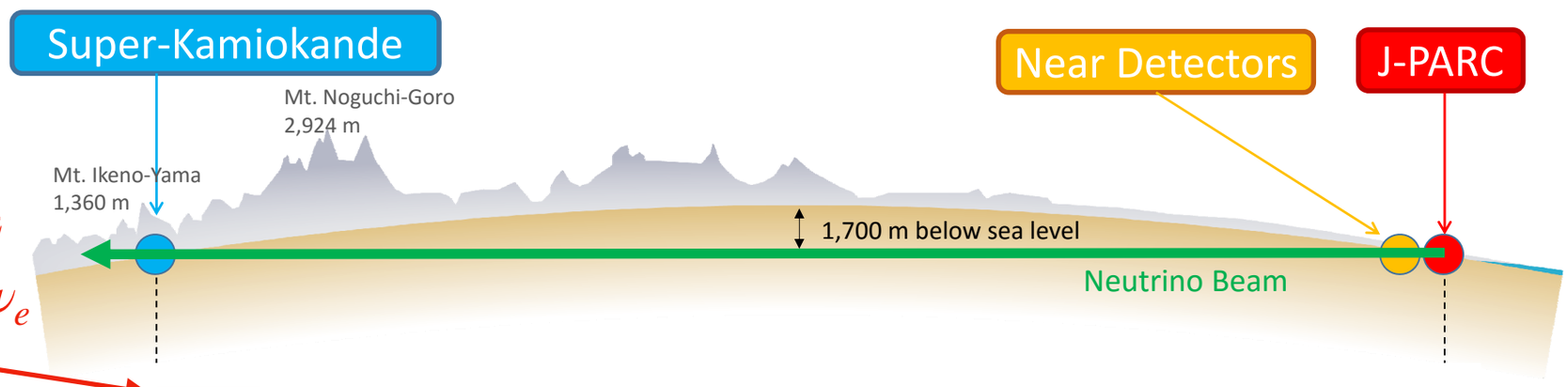
- value of $\delta_{\rm CP}$ → if $\sin\delta_{\rm CP} \neq 0$, CP violation
- sign of Δm^2_{32} (mass ordering)
- is θ_{23} maximal? octant? (i.e. $\theta_{23} < \frac{\pi}{4}$ or $\theta_{23} > \frac{\pi}{4}$)

important params for cosmology (leptogenesis...) and $0\nu2\beta$ searches

flavor symmetries?



Credit: JUNO Collaboration / JGU-Mainz



For $\delta_{\rm CP}$ look for $\nu/\bar{\nu}$ difference of $\nu_\mu \to \nu_e$ appearance

note: wrong-sign ν_μ and intrinsic ν_e backgrounds neglected for illustration

Neutrino mode: $\nu_\tau, \nu_\tau, \nu_\tau, \boxed{\nu_e, \nu_e}, \nu_\mu, \nu_\mu \leftarrow \nu_\mu, \nu_\mu, \nu_\mu, \nu_\mu, \nu_\mu, \nu_\mu, \nu_\mu$

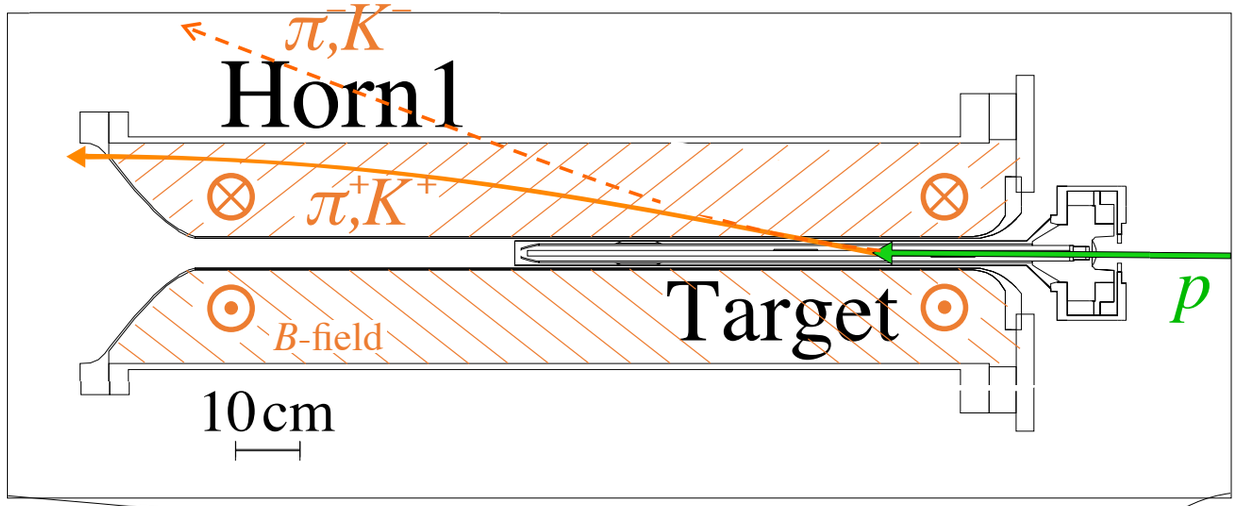
Anti-neutrino mode: $\bar{\nu}_\tau, \bar{\nu}_\tau, \bar{\nu}_\tau, \bar{\nu}_\tau, \boxed{\bar{\nu}_e, \bar{\nu}_\mu}, \bar{\nu}_\mu \leftarrow \bar{\nu}_\mu, \bar{\nu}_\mu, \bar{\nu}_\mu, \bar{\nu}_\mu, \bar{\nu}_\mu, \bar{\nu}_\mu, \bar{\nu}_\mu$

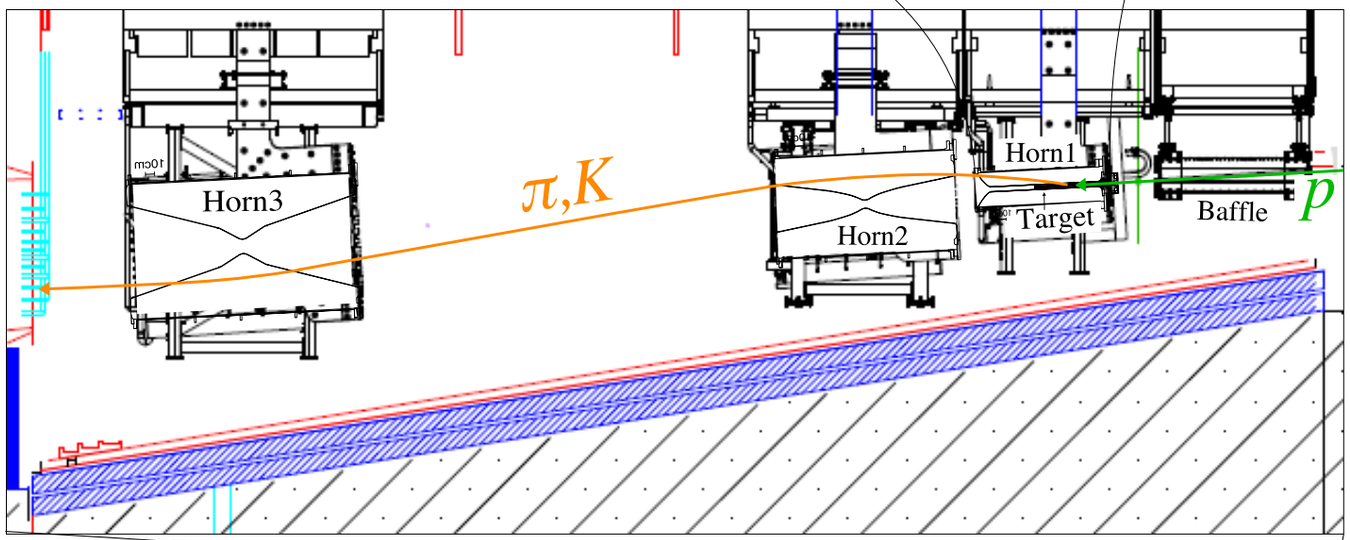
Neutrino beam

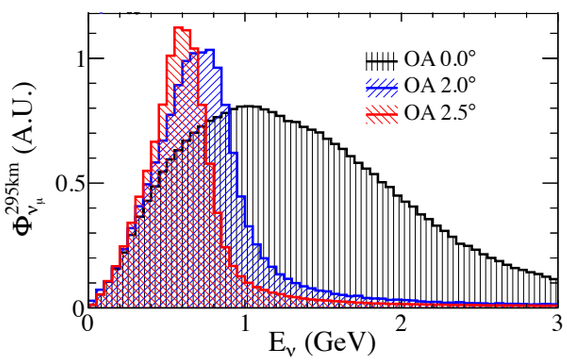
- 30 GeV protons produce π,K in 90 cm graphite target

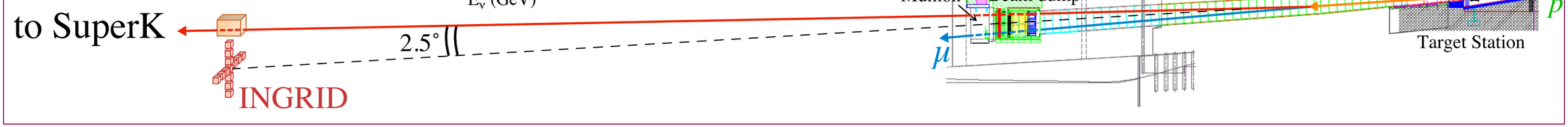
- Three magnetic horns selectively focus π^+,K^+ or π^-,K^- to produce ν_μ or $\bar{\nu}_\mu$ beam (decay in-flight).

- Narrowband beam thanks to off-axis technique.

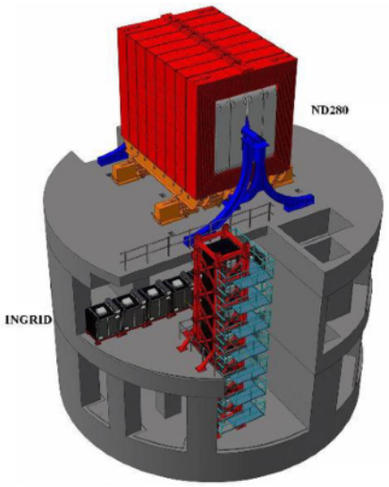


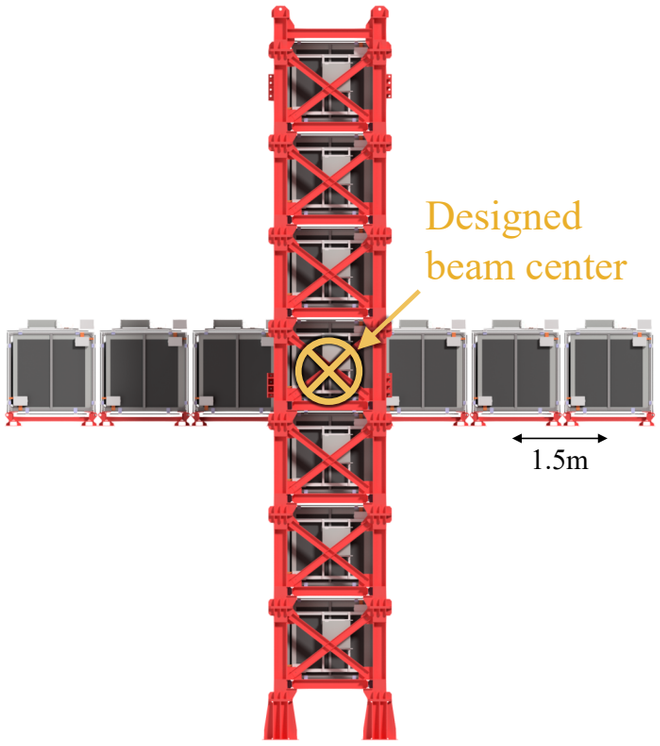


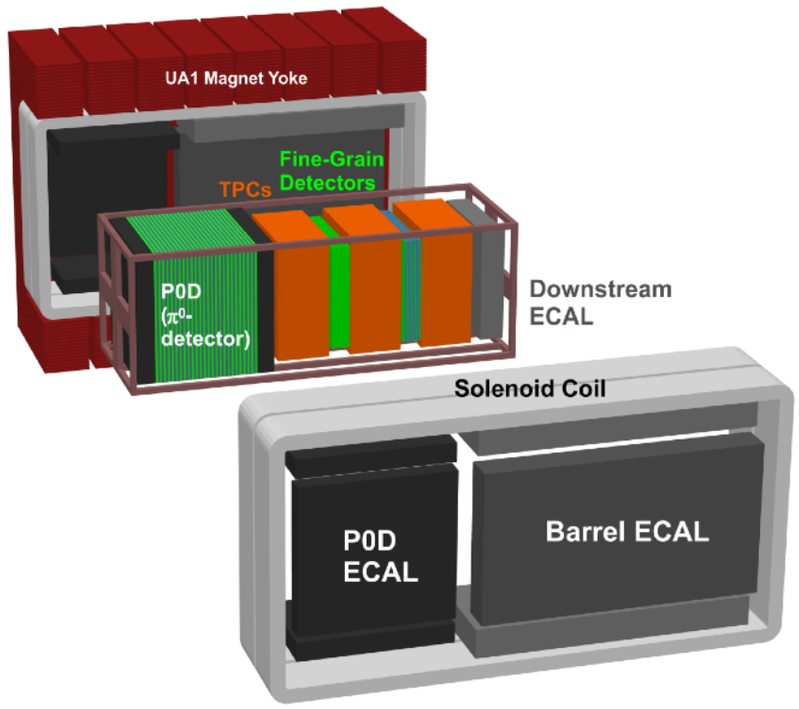


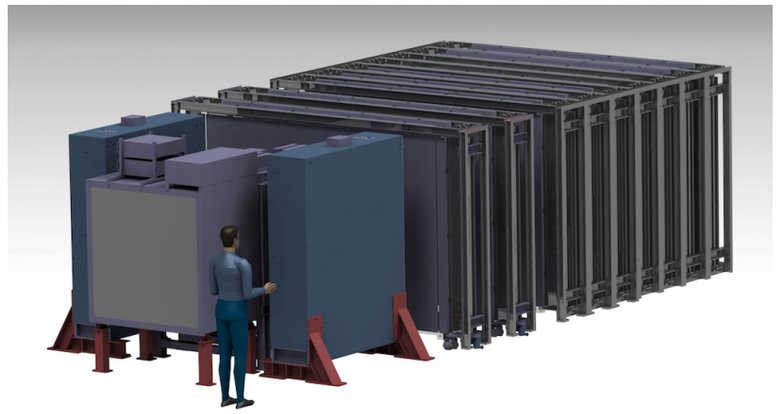


The near detectors









INGRID on-axis detector

- Iron-scintillator sandwich detectors monitor neutrino beam **direction** and intensity

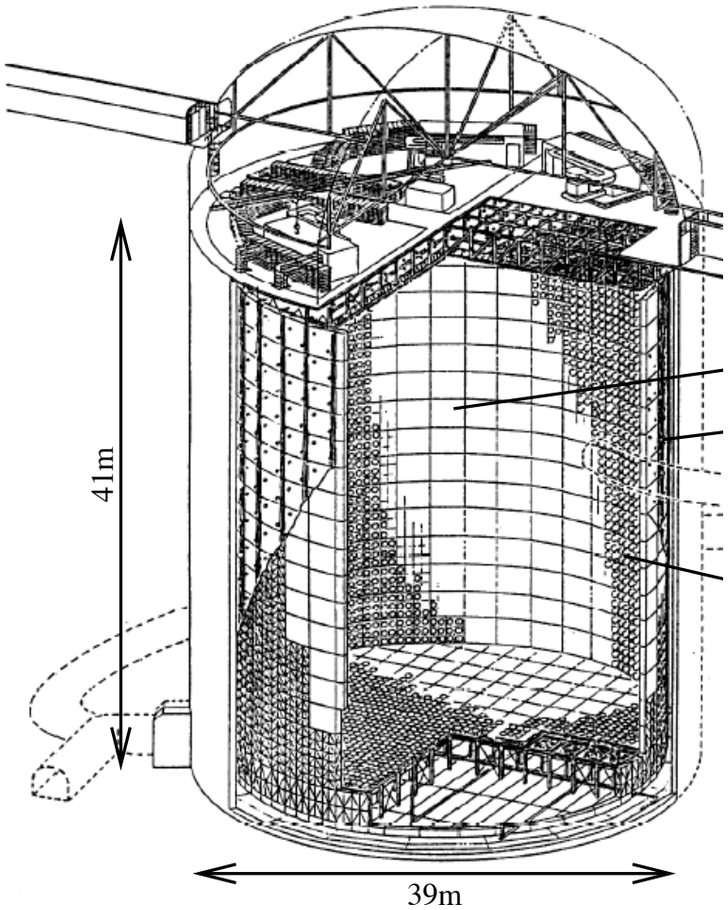
ND280 off-axis detector

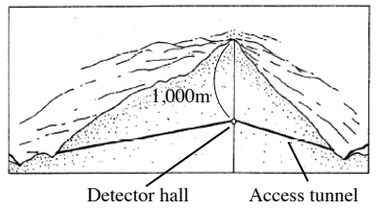
- Active **scintillator** + passive **water** targets
- Tracking with **time projection chambers**
- **Magnetized** for charge and momentum measurement

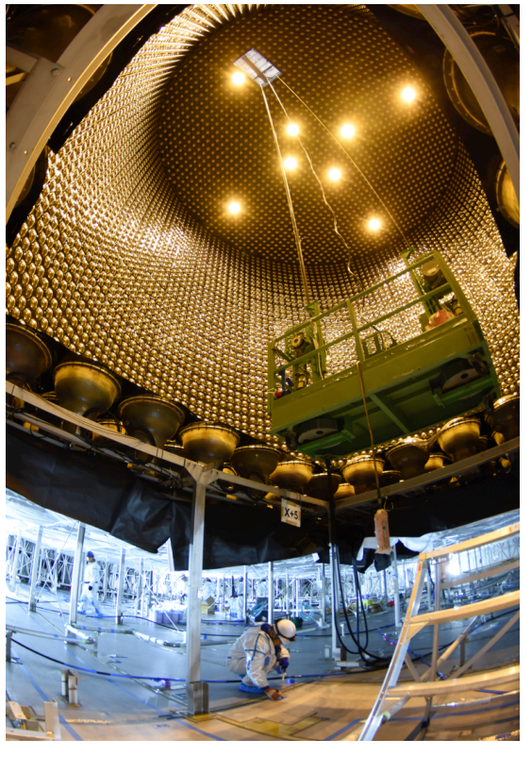
WAGASCI + BabyMIND

- Latest addition at intermediate **1.5º** off-axis flux
- **Water** target with cuboid lattice scintillators for high angle acceptance
- Compact **magnetized** iron muon range detector
- First xsec meas. published: PTEP, ptab014 (2021) new!

SuperK – the far detector



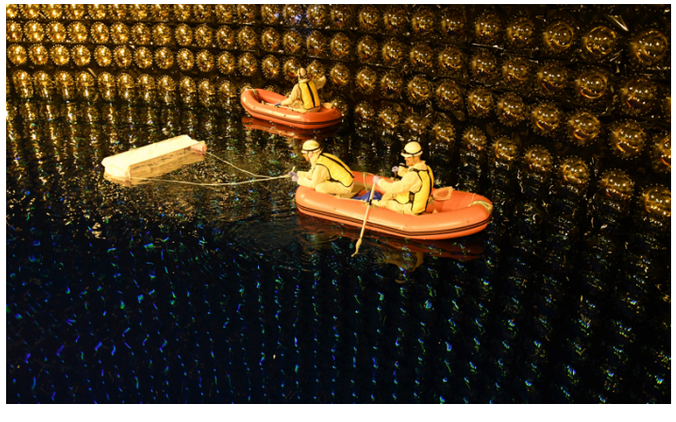




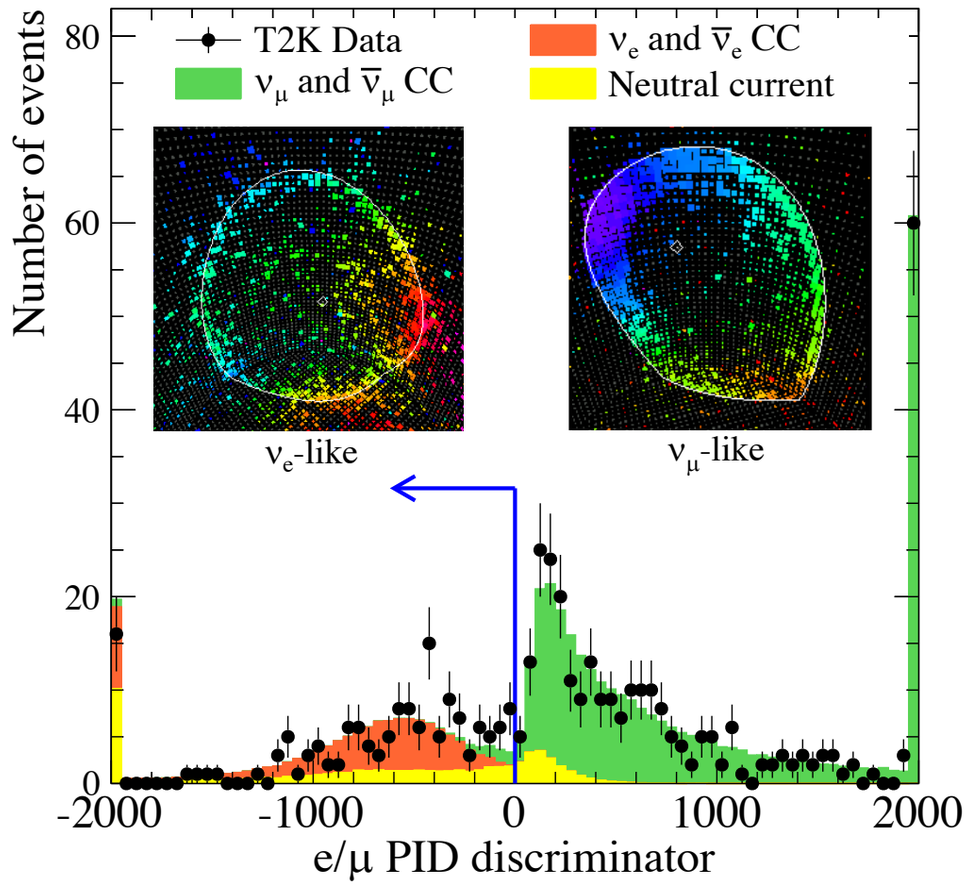
50 kt pure water
~11,000 20 inch PMTs

Gadolinium loading has started for improved neutron tagging!

Photos: "Super Kamiokande refurbishment" ICRR (2018)



Adapted from Nature 580, 339-344 (2020)



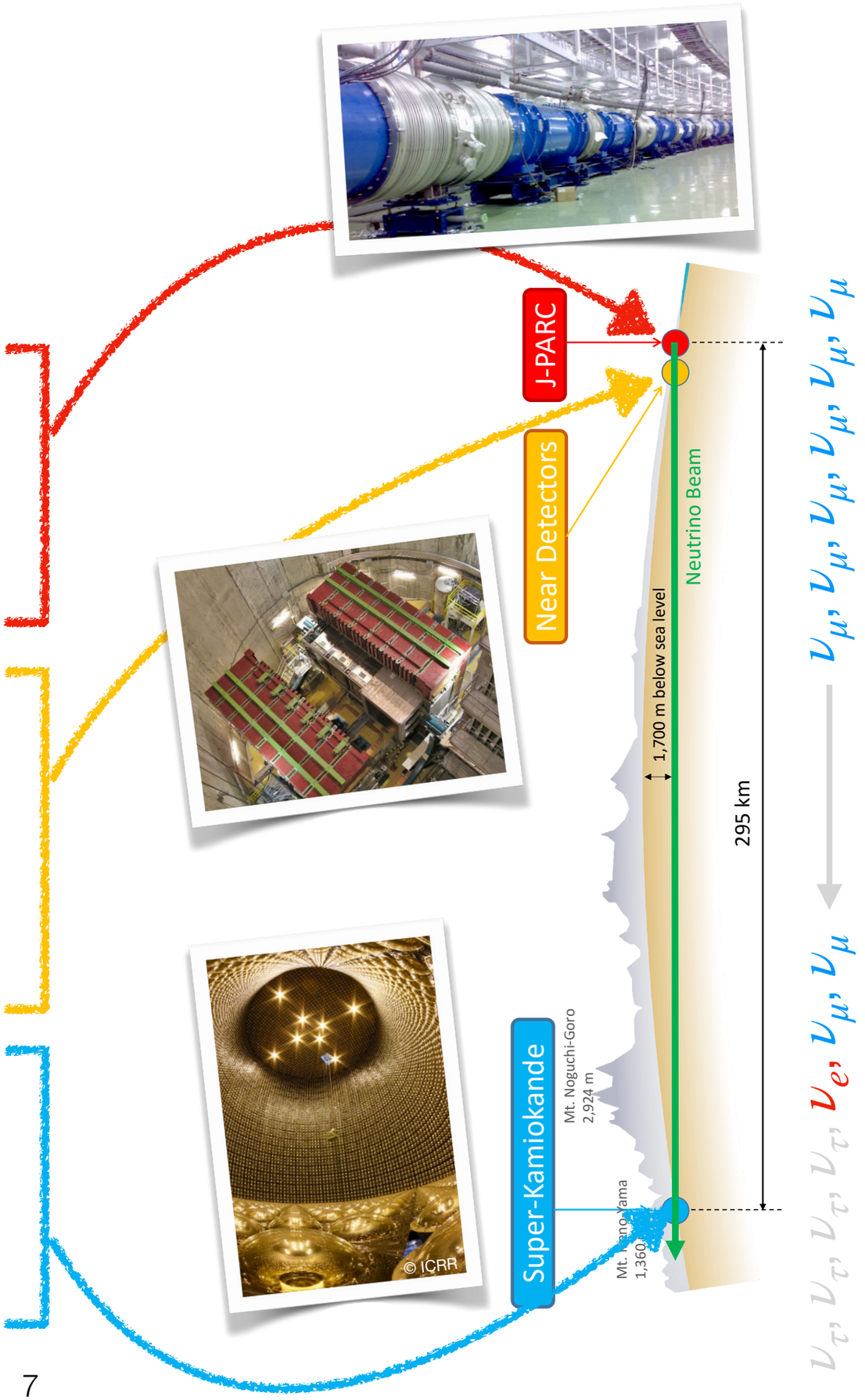
- Good μ/e PID from ring shape
- Reconstruct neutrino energy from lepton momentum and angle w.r.t. neutrino beam
- Not magnetized, so the beam $\nu/\bar{\nu}$-modes are important. ND280 further constrains the wrong-sign background.

Analysis strategy

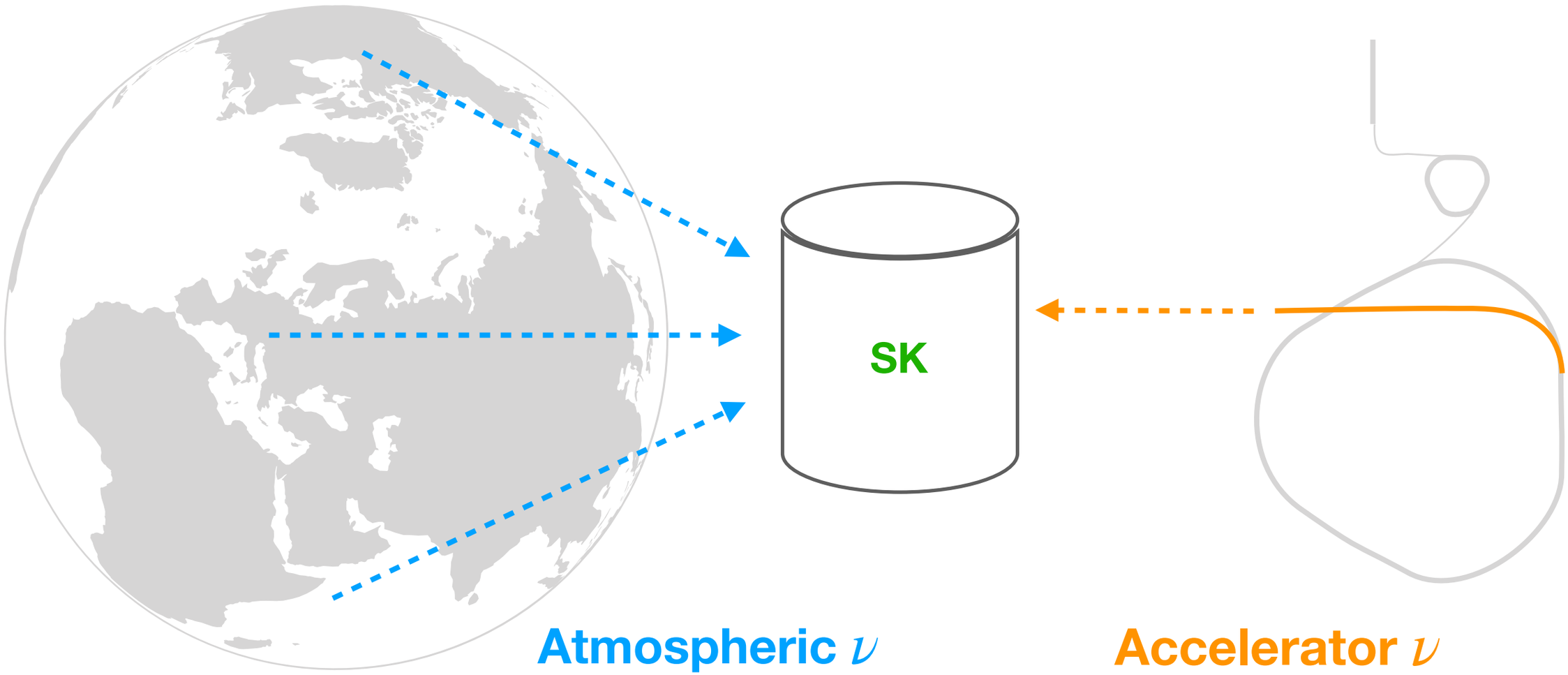
- Beam monitors + hadron production experiments
  → **neutrino flux**

- ND280 measurements
  + interaction model
  + external constraints
  → **unoscillated flux × xsec**

- 5 samples at SK
  → ν_μ **disappearance +**
  ν_e **appearance**

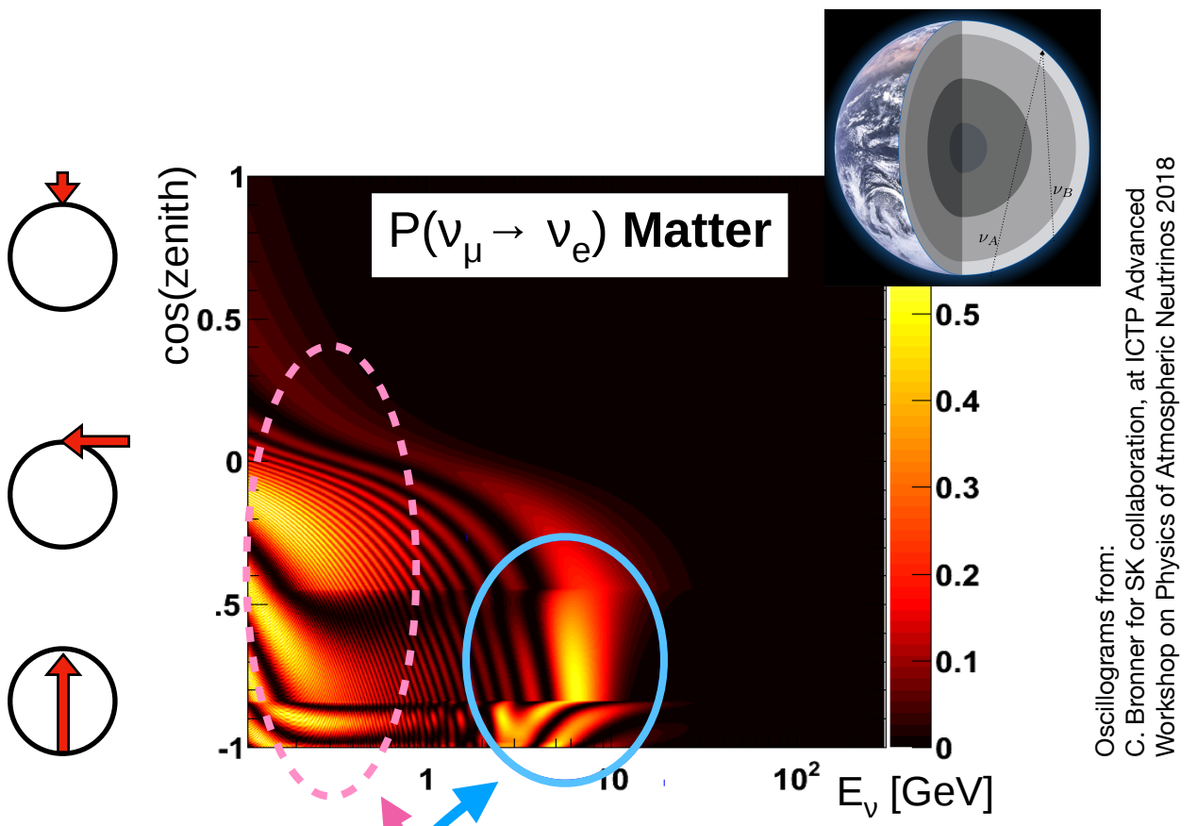


Thesis: a joint fit between SK atmospheric and T2K accelerator ν



CP and mass ordering sensitivity

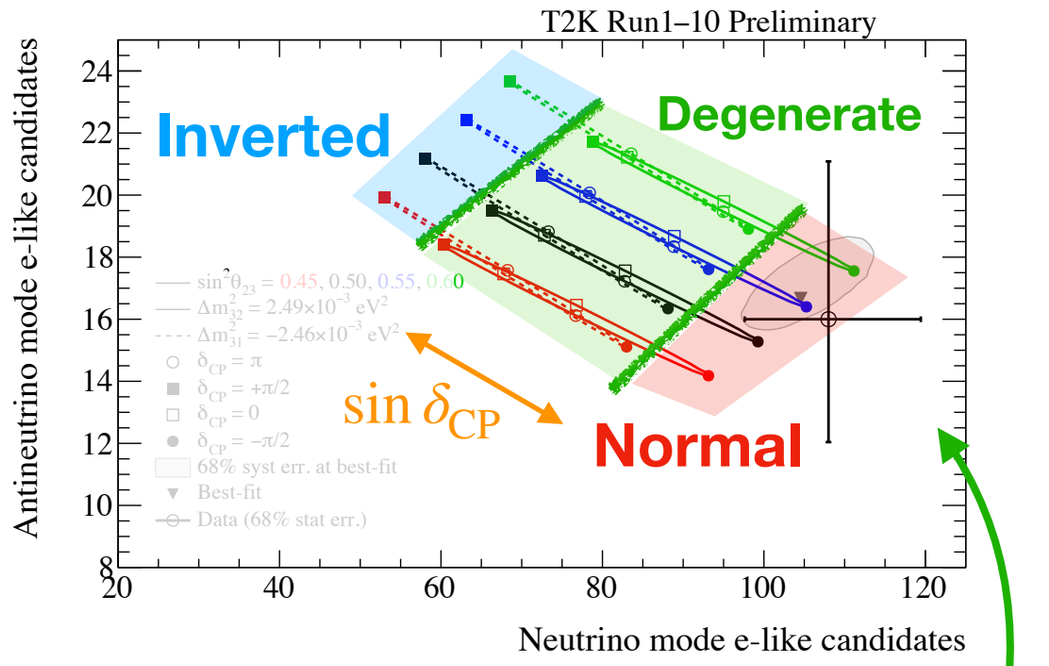
SK Atmospheric



Oscillograms from:
C. Bronner for SK collaboration, at ICTP Advanced Workshop on Physics of Atmospheric Neutrinos 2018

- Resonance in Earth mantle & core sensitive to mass ordering
- Weakly sensitive to δ_{CP} via normalization of sub-GeV e-like

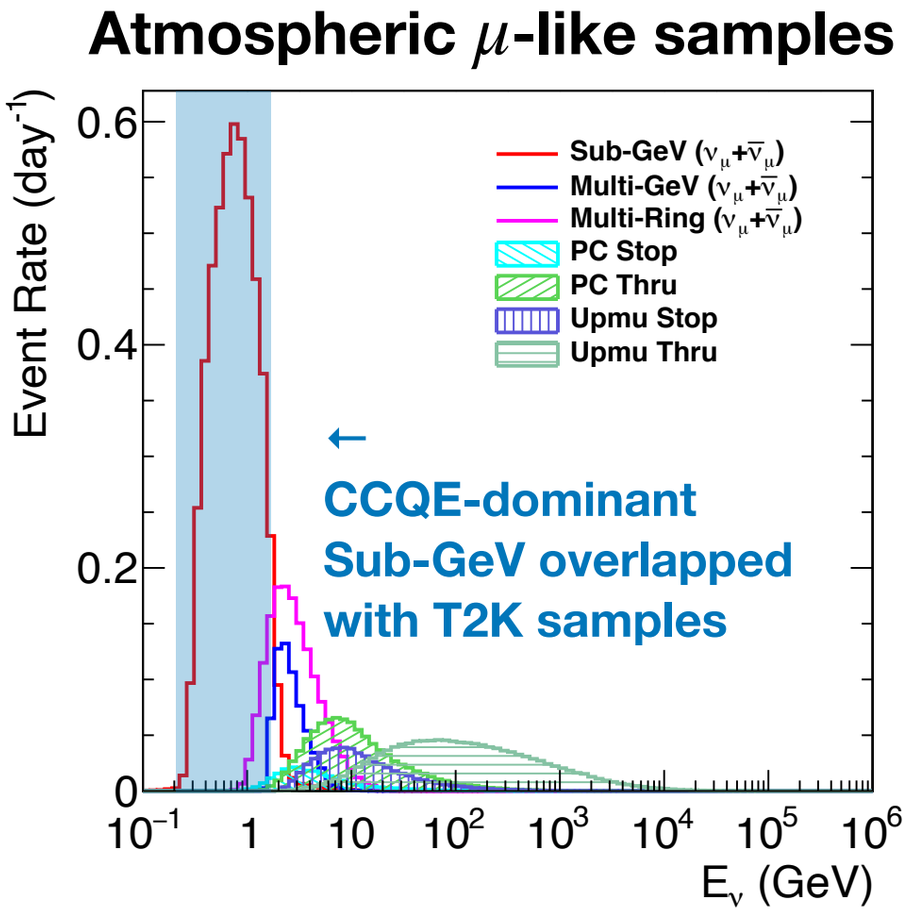
T2K Accelerator



- Anti-correlated change of $\nu_e, \bar{\nu}_e$ appearance probability → δ_{CP}
- For large changes also weakly sensitive to mass ordering

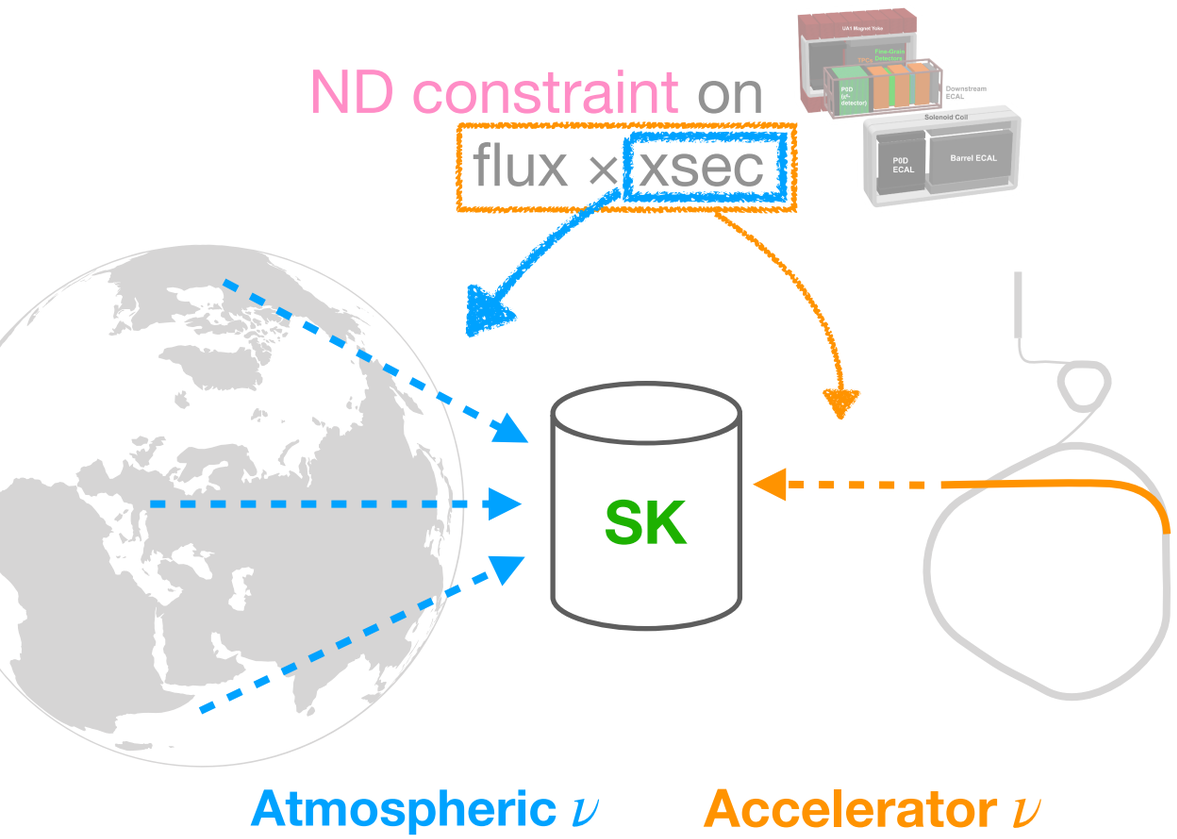
degenerate

When combined, can resolve degeneracy and have better CP violation sensitivity!



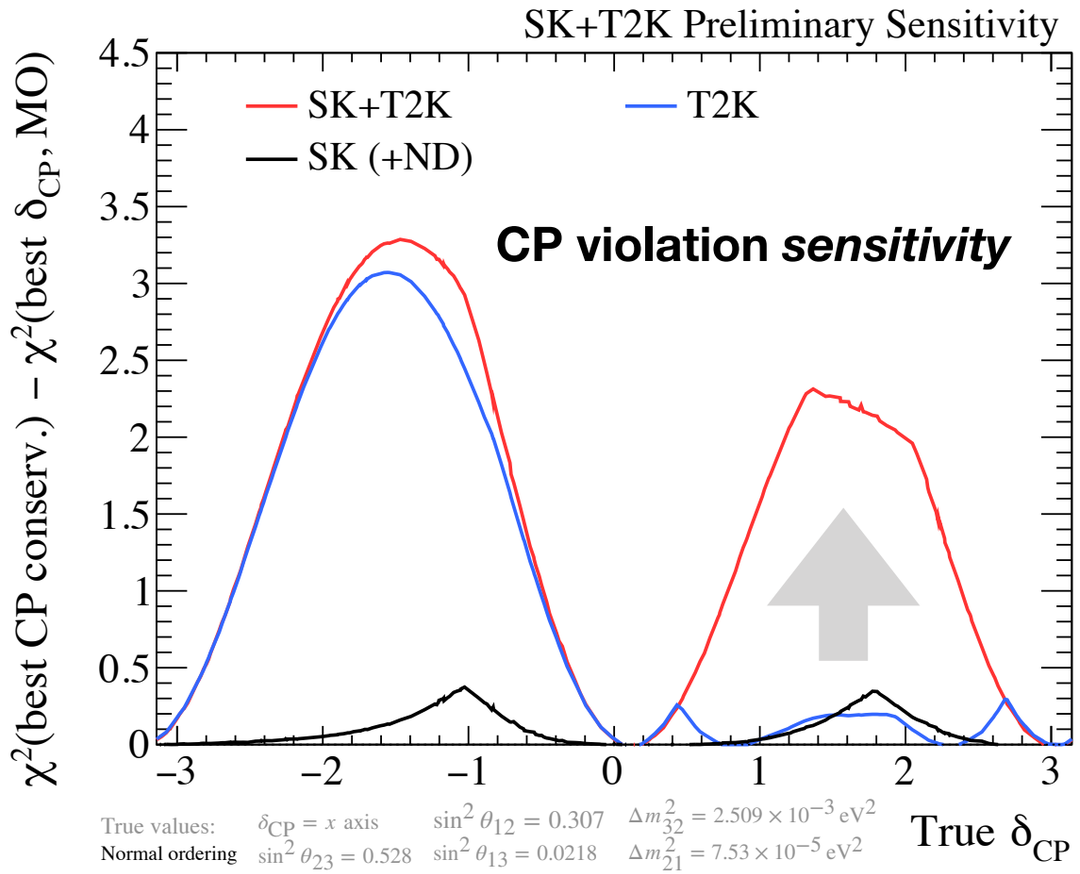
Systematic correlations

- Overlapped true energy region
  → coherent **interaction model** to capture correlations
  → Bonus: **ND constraint** for atmospherics!



- Same Super-K detector used by both experiments
  → estimate contribution from **detector syst.** correlations

Sensitivities



- Both experiments complement each other in many parameters,

  effect of degeneracy-resolution is quite nice

- At current statistics, contribution from systematic correlations on sensitivity seems limited

- This kind of analysis very important for future HyperK experiment

from Neutrino 2022

Efficient statistics for future neutrino oscillation experiments

personal work

hopefully useful for a wider group than just our experiment

Inference

- Toy example: threw coin $n = 100$ times, got heads $x = 40$ times,

  what is the probability θ of this coin to give heads?

- Probability distribution to generate the data

  $$p(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

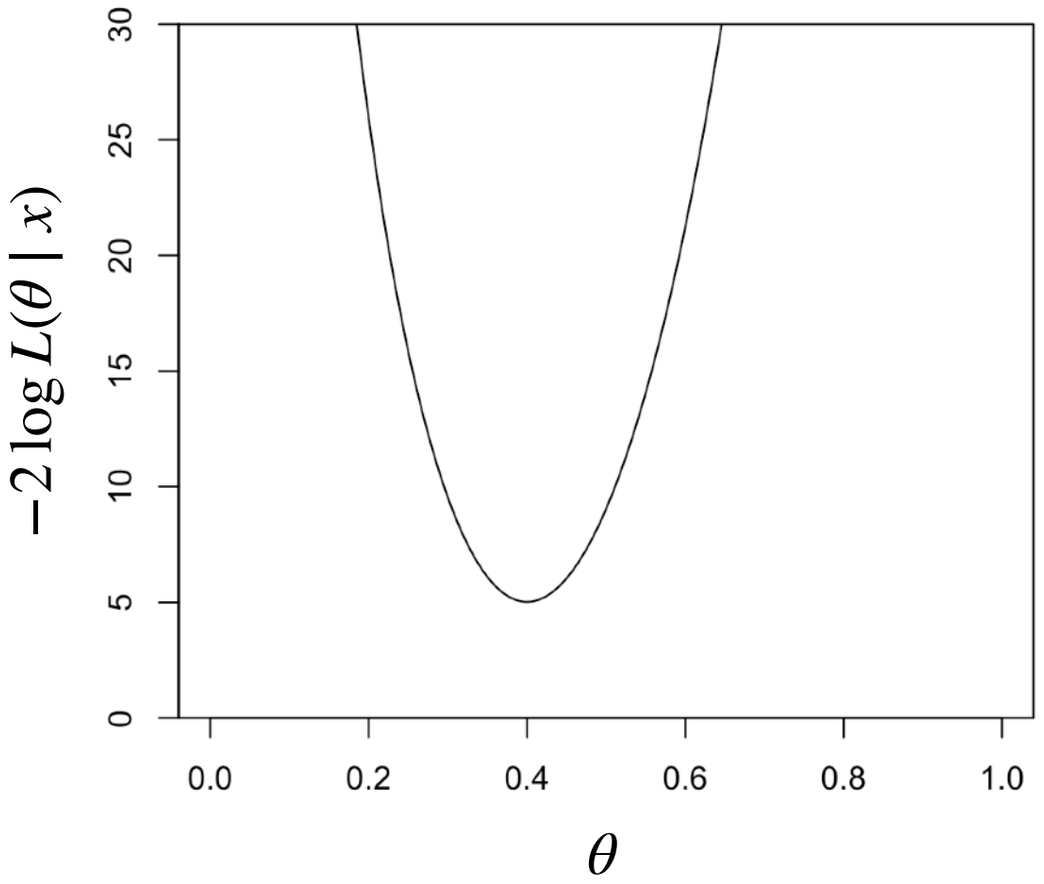
  seen as function of θ this is called the likelihood

  $$L(\theta \mid x) := p(x \mid \theta)$$

- Best agreement = maximum likelihood at

  $$\hat{\theta} = x/n$$

  better: interval over θ values to account for statistical fluctuations



Confidence intervals

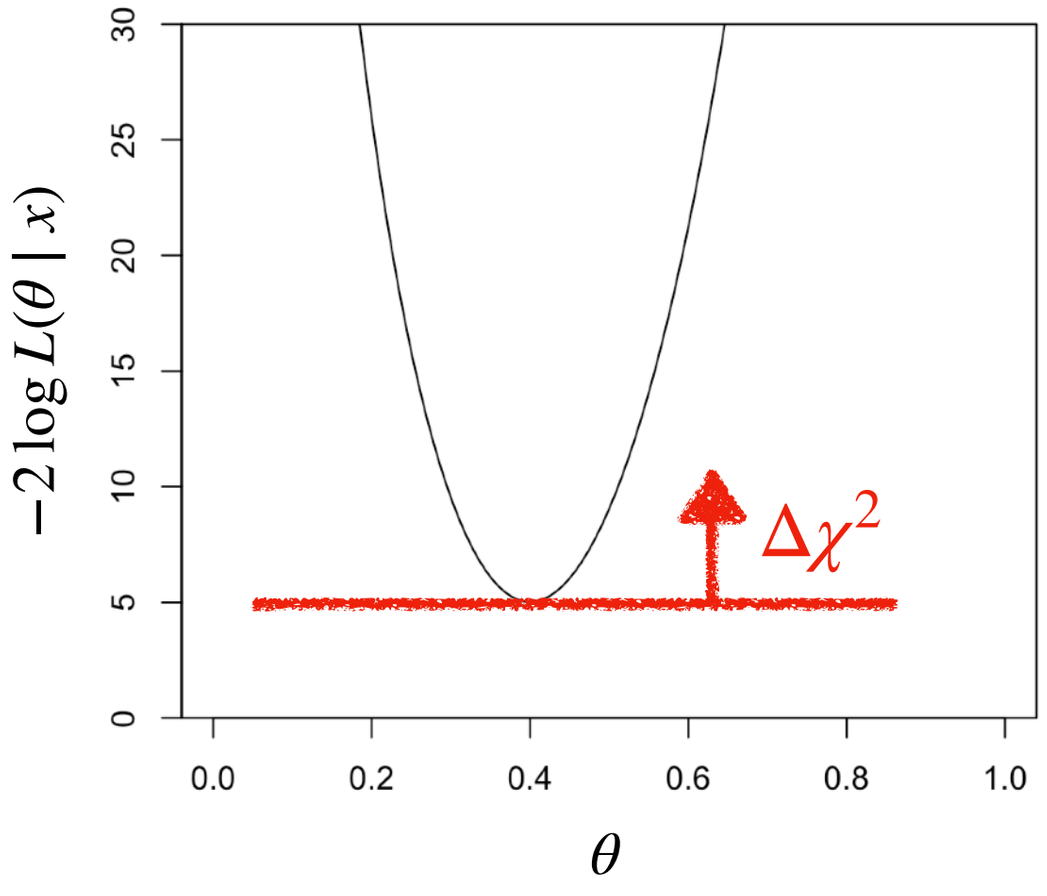
- Confidence interval:
  Whatever the true value is, the interval would cover it with at least 68%, 90%, …

- Define $\chi^2(\theta) := -2 \log L(\theta \mid x)$.
  Small $\chi^2(\theta)$ in general means better agreement with data.

- Study how much worse than at the minimum we are: $\Delta\chi^2(\theta) := \chi^2(\theta) - \chi^2_{\text{min}}$

- Strategy: create a confidence interval using $\Delta\chi^2$

  Choose range of θ such that $\Delta\chi^2(\theta) < \Delta\chi^2_c$
  with some critical value $\Delta\chi^2_c$



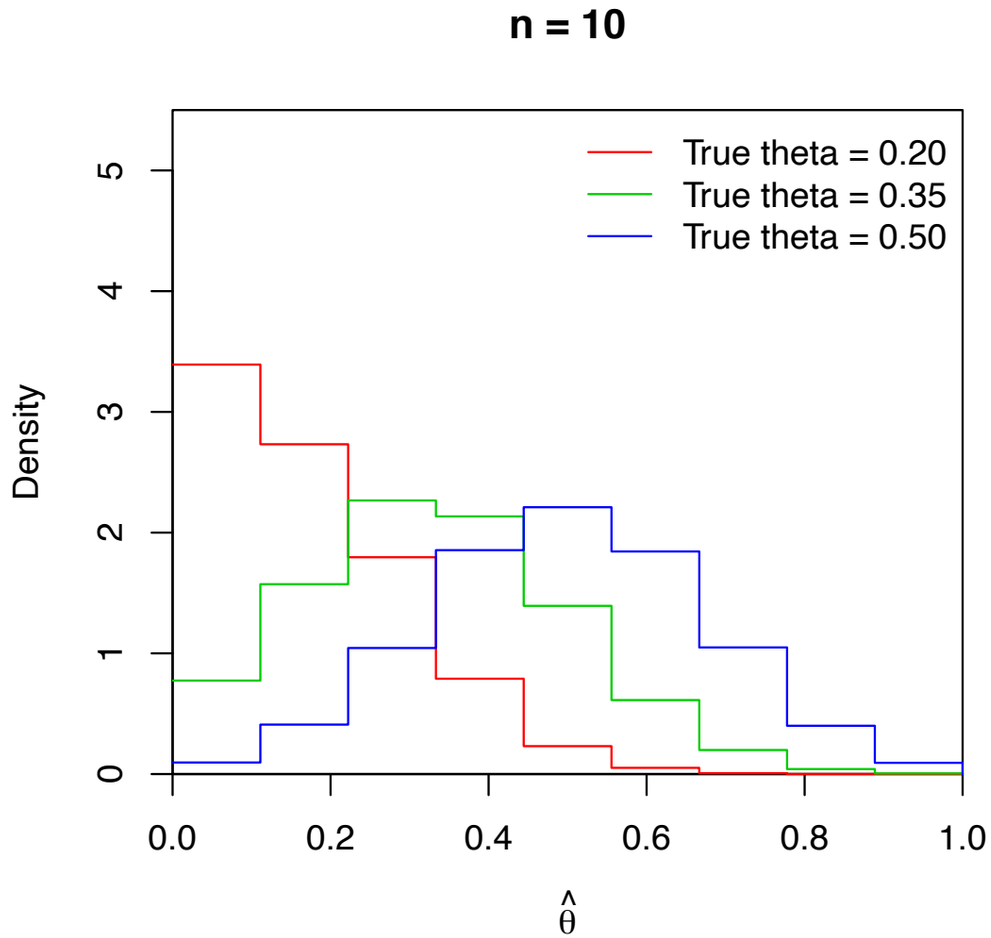
- In the presence of nuisance parameters η, choose e.g. the best agreement

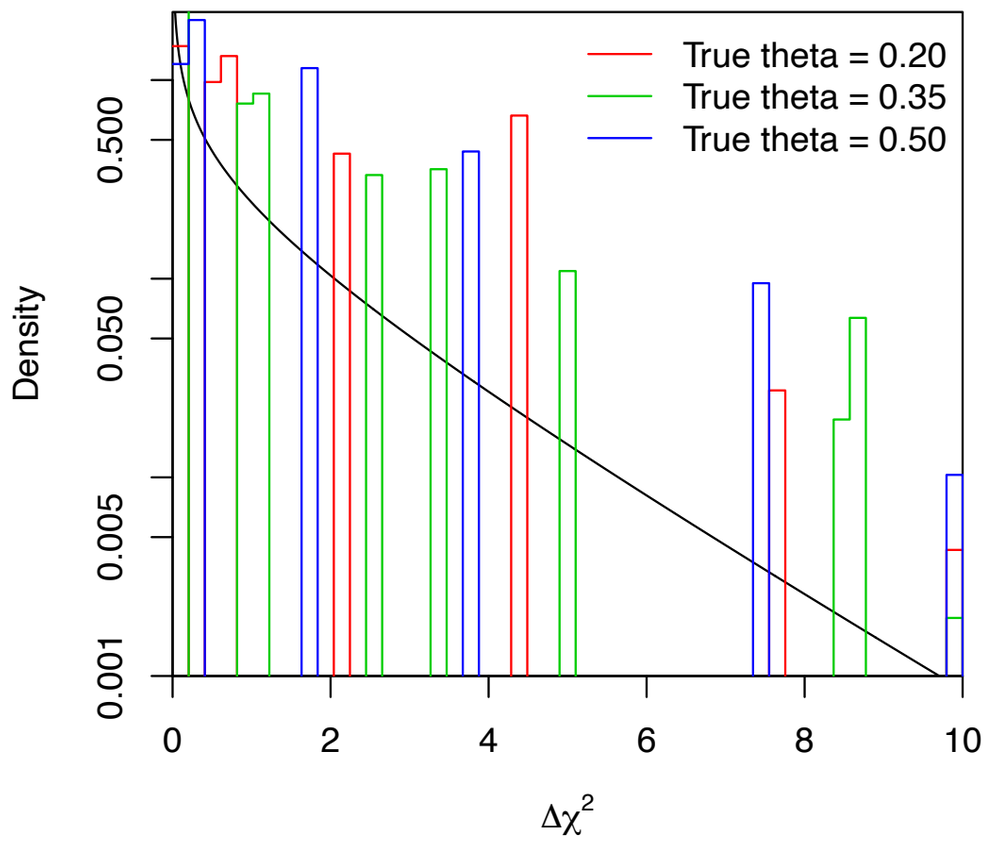
$$\chi^2_p(\theta) = \min_{\eta} \chi^2(\theta, \eta)$$

Wilk's theorem

- Under certain conditions
  - regardless of $p(x \mid \theta)$
  - regardless of θ_{true}
    (in the case of profiling also regardless of true nuisance η_{true})
  - regardless of n
  - in $n \to \infty$ limit, $\Delta\chi^2$ evaluated at the true value
    becomes a chi-squared distribution,
    i.e. $p\left(\Delta\chi^2(\theta_{\text{true}})\right) = p(\chi^2_k)$ where k is dimension of θ

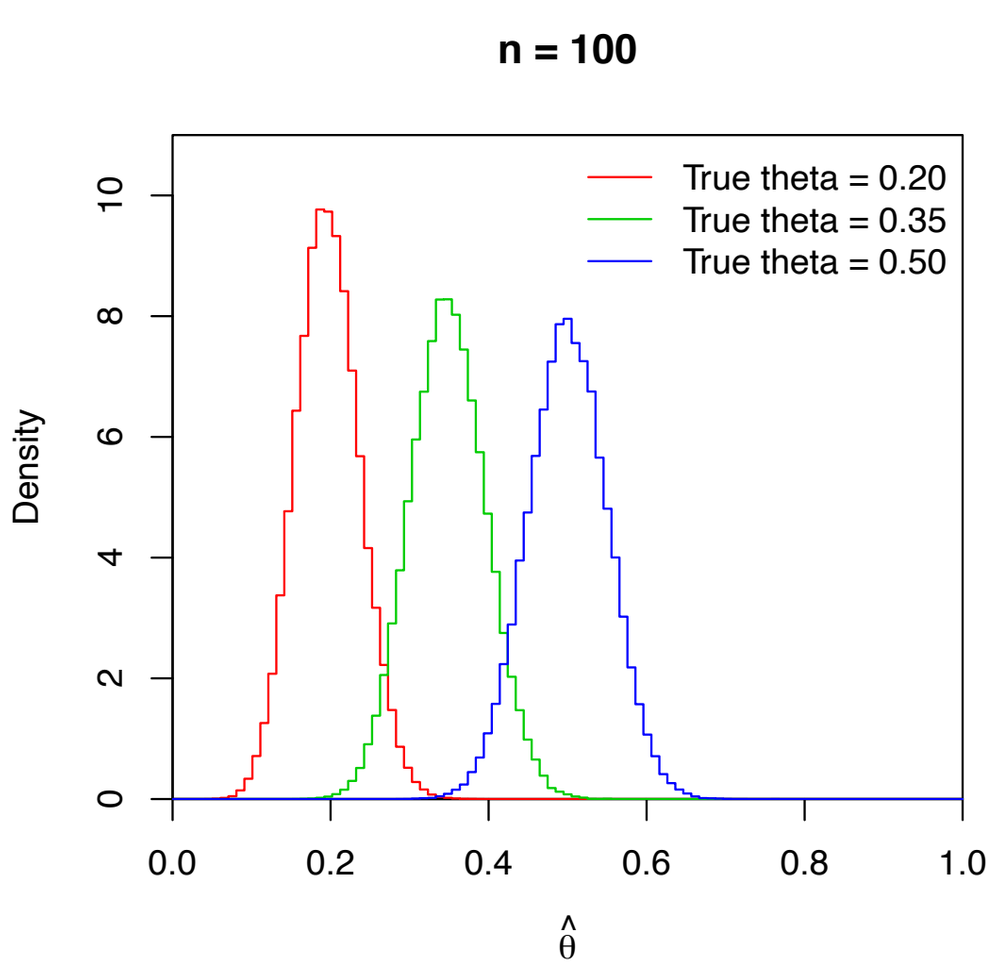
- So can just construct interval by choosing
  e.g. $\Delta\chi^2(\theta) < 1$ for a 68% interval if $k = 1$.

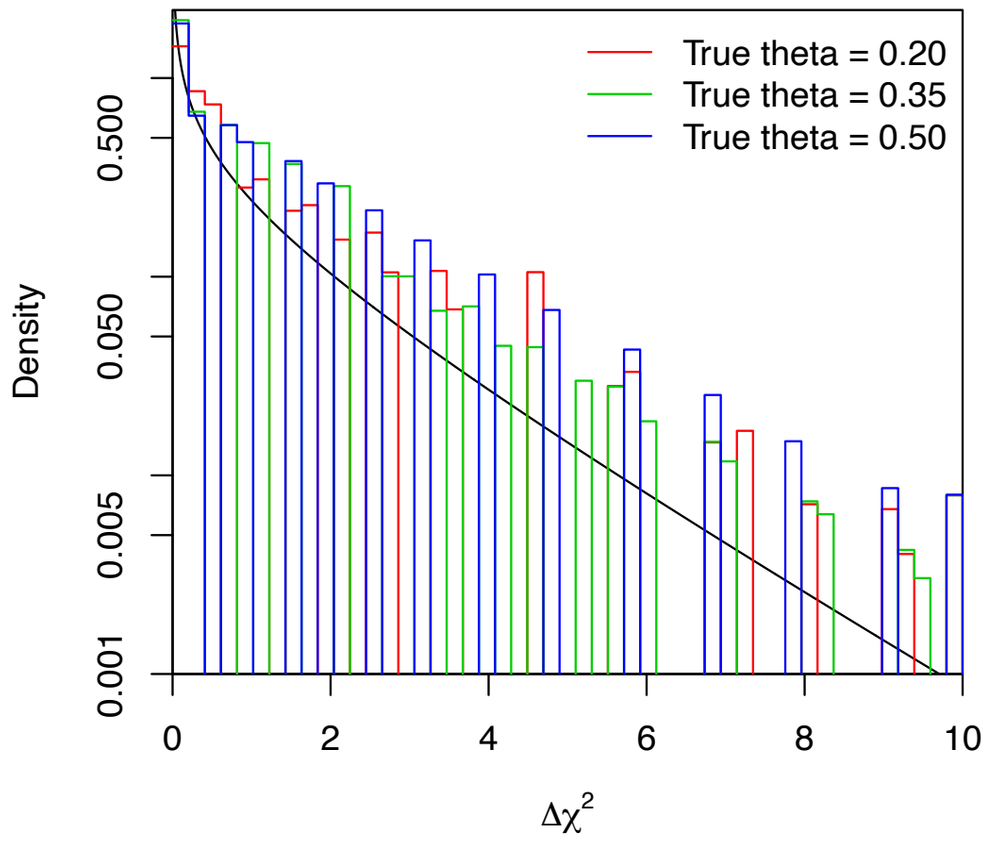




Wilk's theorem

- Under certain conditions
  - regardless of $p(x \mid \theta)$
  - regardless of θ_{true}
    (in the case of profiling also regardless of true nuisance η_{true})
  - regardless of n
  - in $n \to \infty$ limit, $\Delta\chi^2$ evaluated at the true value
    becomes a chi-squared distribution,
    i.e. $p\left(\Delta\chi^2(\theta_{\text{true}})\right) = p(\chi^2_k)$ where k is dimension of θ
- So can just construct interval by choosing
  e.g. $\Delta\chi^2(\theta) < 1$ for a 68% interval if $k = 1$.

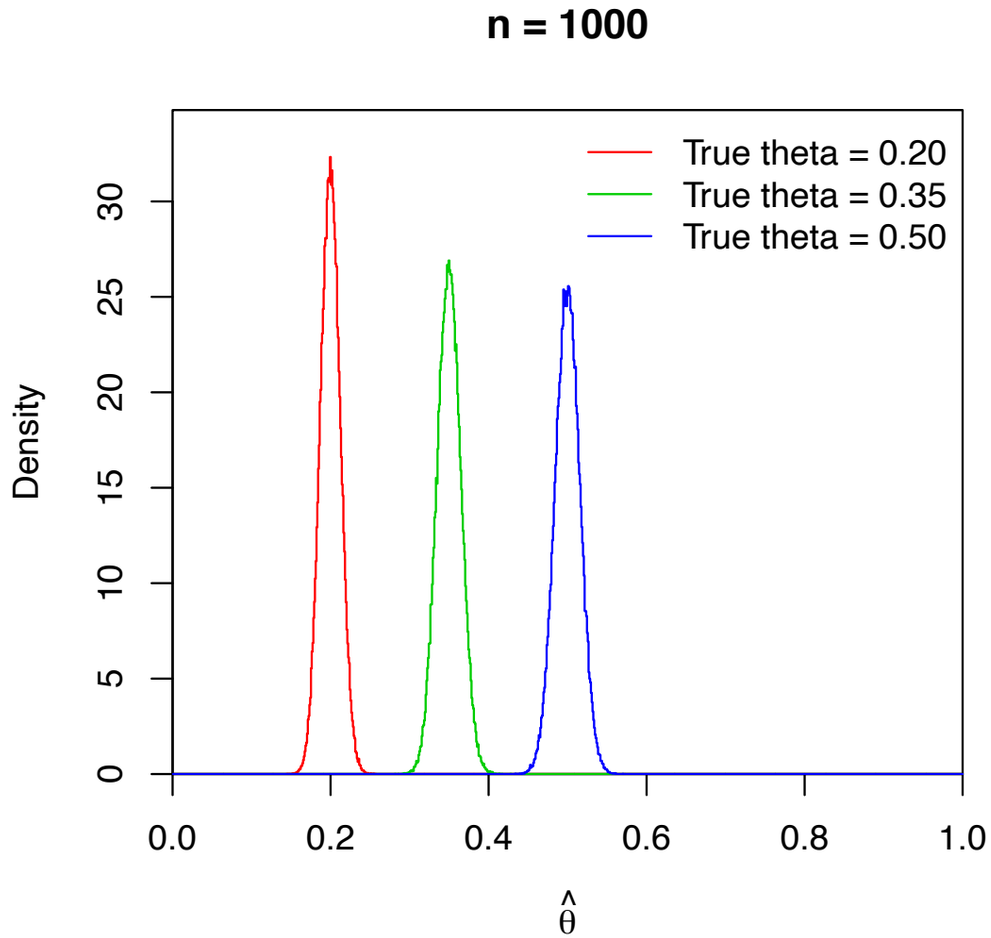


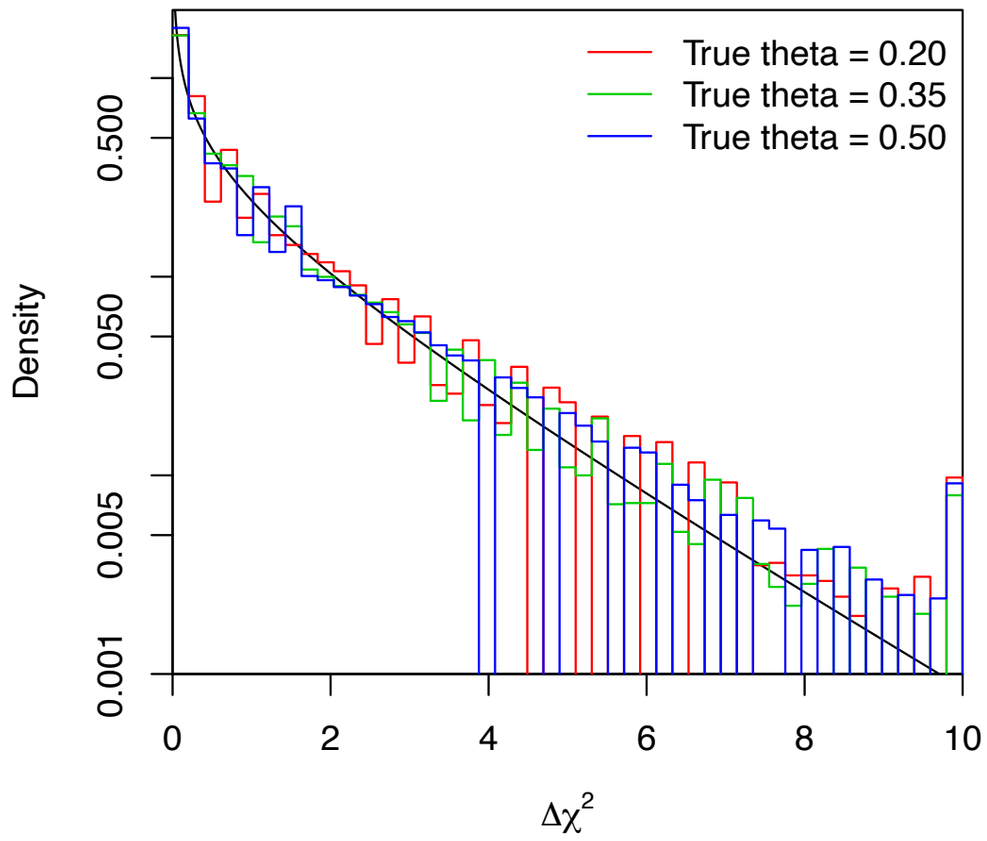


Wilk's theorem

- Under certain conditions
  - regardless of $p(x \mid \theta)$
  - regardless of θ_{true}
    (in the case of profiling also regardless of true nuisance η_{true})
  - regardless of n
  - in $n \to \infty$ limit, $\Delta\chi^2$ evaluated at the true value
    becomes a chi-squared distribution,
    i.e. $p\left(\Delta\chi^2(\theta_{\text{true}})\right) = p(\chi^2_k)$ where k is dimension of θ

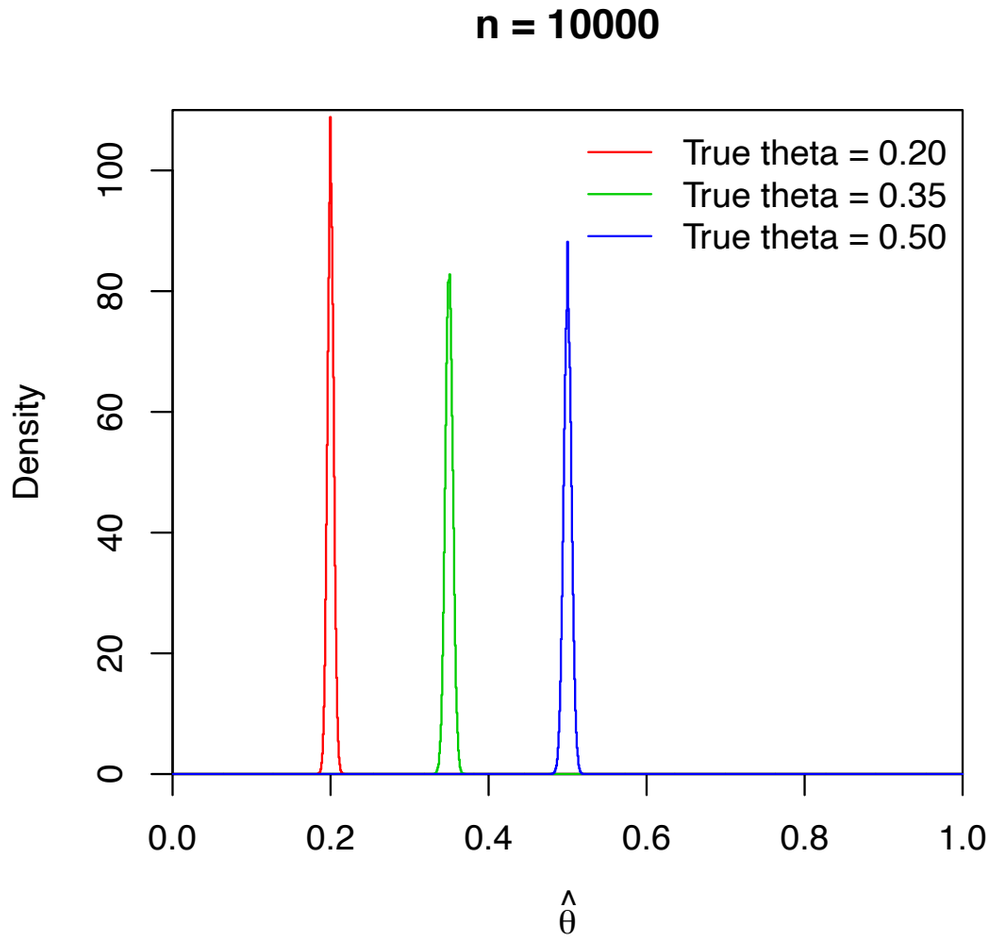
- So can just construct interval by choosing
  e.g. $\Delta\chi^2(\theta) < 1$ for a 68% interval if $k = 1$.

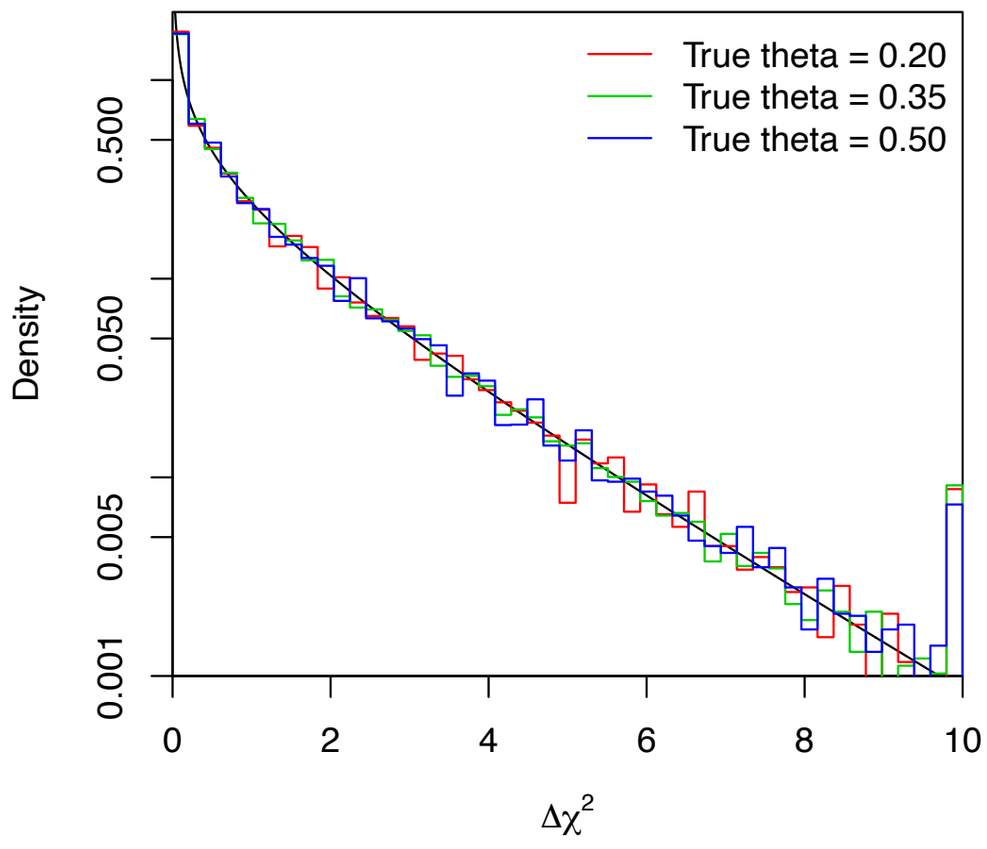




Wilk's theorem

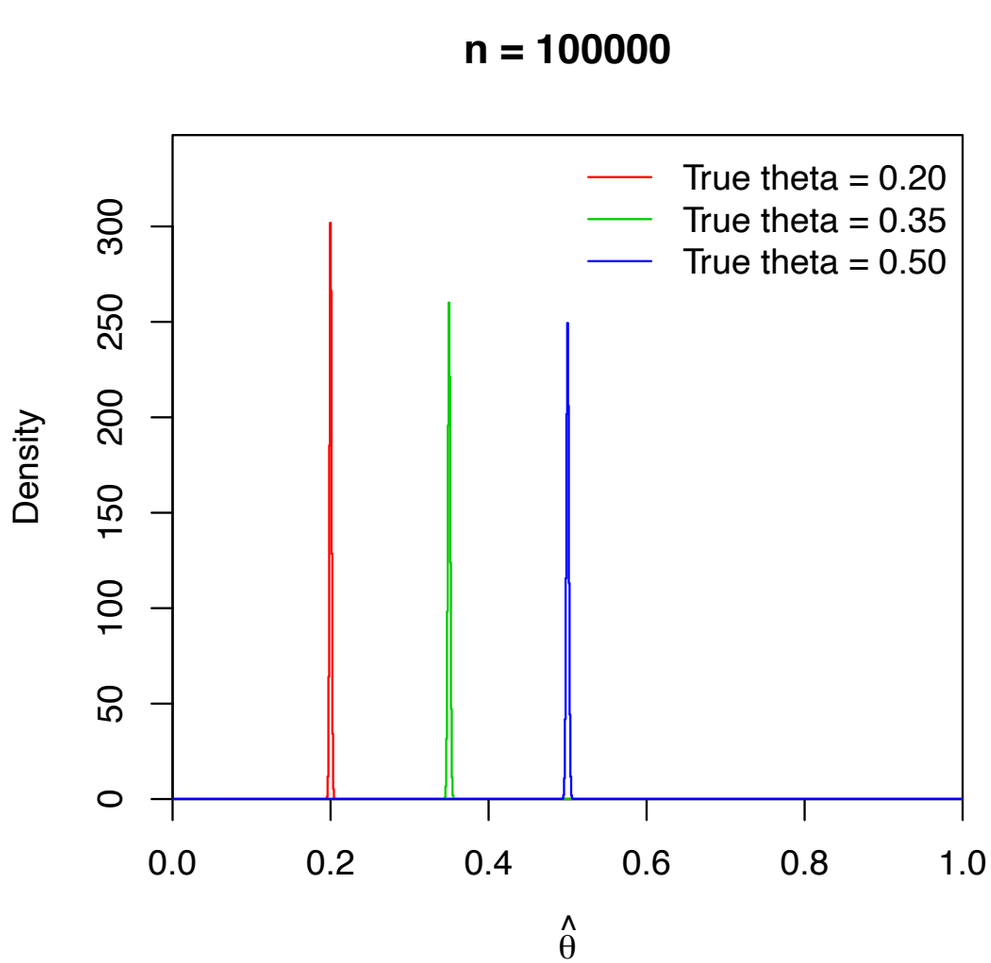
- Under certain conditions
  - regardless of $p(x \mid \theta)$
  - regardless of θ_{true}
    (in the case of profiling also regardless of true nuisance η_{true})
  - regardless of n
  - in $n \to \infty$ limit, $\Delta\chi^2$ evaluated at the true value
    becomes a chi-squared distribution,
    i.e. $p\left(\Delta\chi^2(\theta_{\text{true}})\right) = p(\chi^2_k)$ where k is dimension of θ
- So can just construct interval by choosing
  e.g. $\Delta\chi^2(\theta) < 1$ for a 68% interval if $k = 1$.

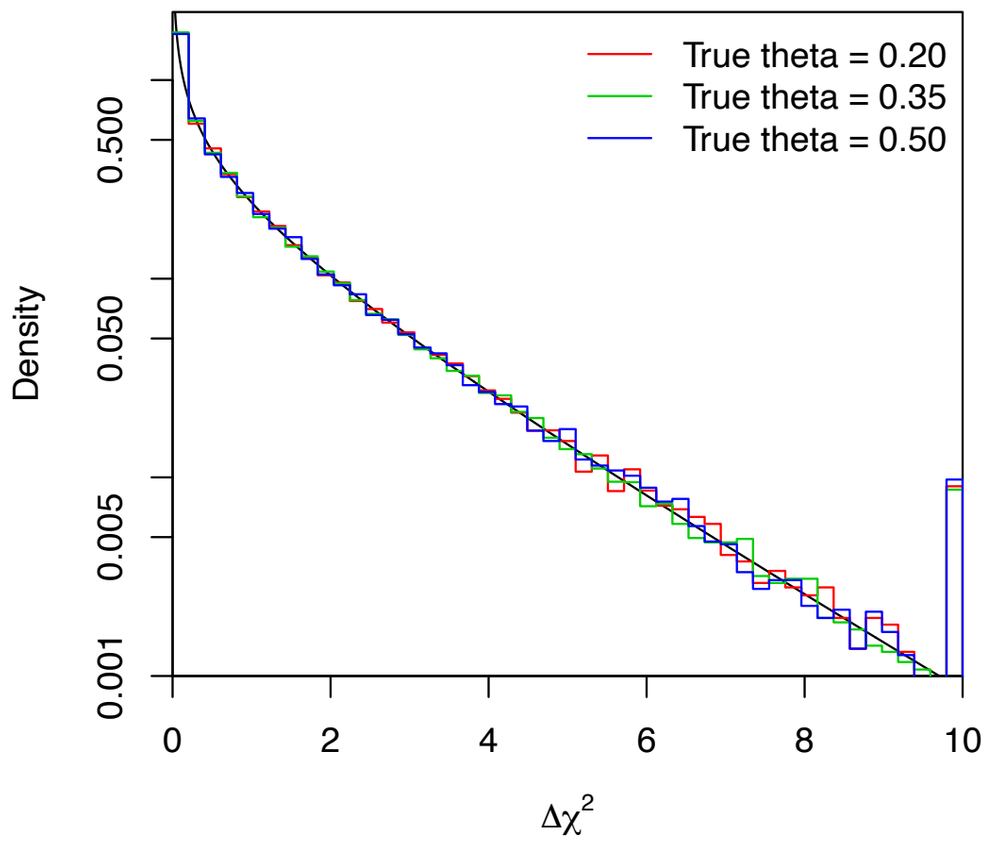




Wilk's theorem

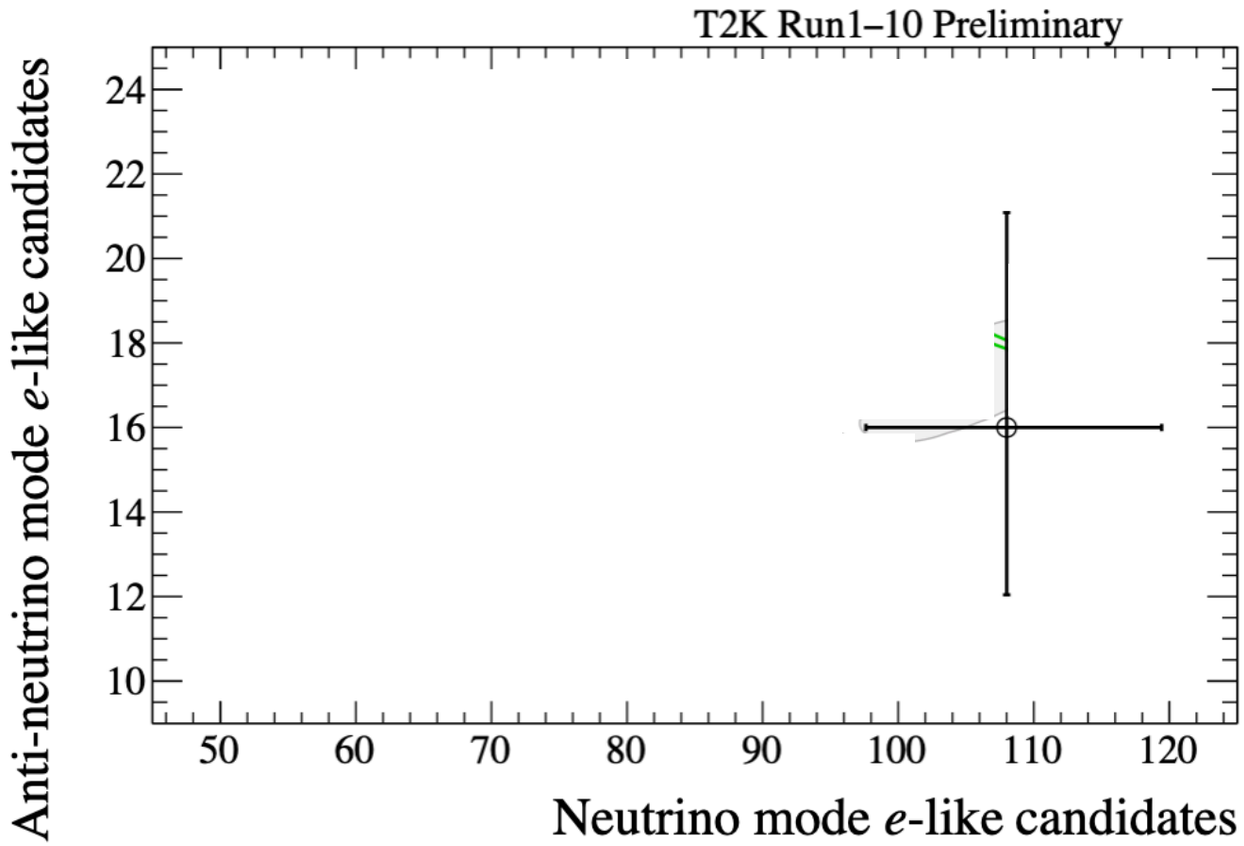
- Under certain conditions
  - regardless of $p(x \mid \theta)$
  - regardless of θ_{true}
    (in the case of profiling also regardless of true nuisance η_{true})
  - regardless of n
  - in $n \to \infty$ limit, $\Delta\chi^2$ evaluated at the true value
    becomes a chi-squared distribution,
    i.e. $p\left(\Delta\chi^2(\theta_{\text{true}})\right) = p(\chi^2_k)$ where k is dimension of θ
- So can just construct interval by choosing
  e.g. $\Delta\chi^2(\theta) < 1$ for a 68% interval if $k = 1$.





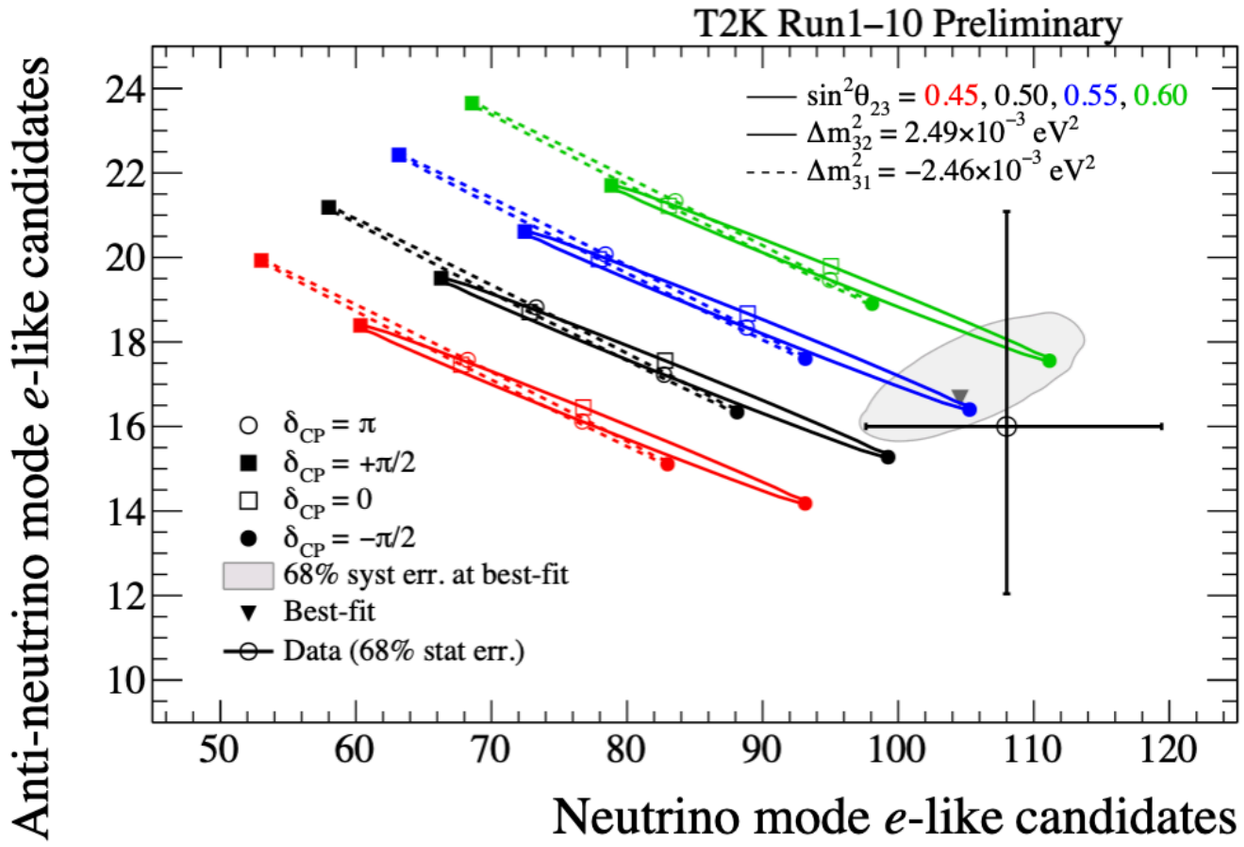
Can we use Wilk's theorem for leptonic CP violation search?

- Small amount of data → no 😭
  ν-mode: 119, $\bar{\nu}$-mode: 16



Can we use Wilk's theorem for leptonic CP violation search?

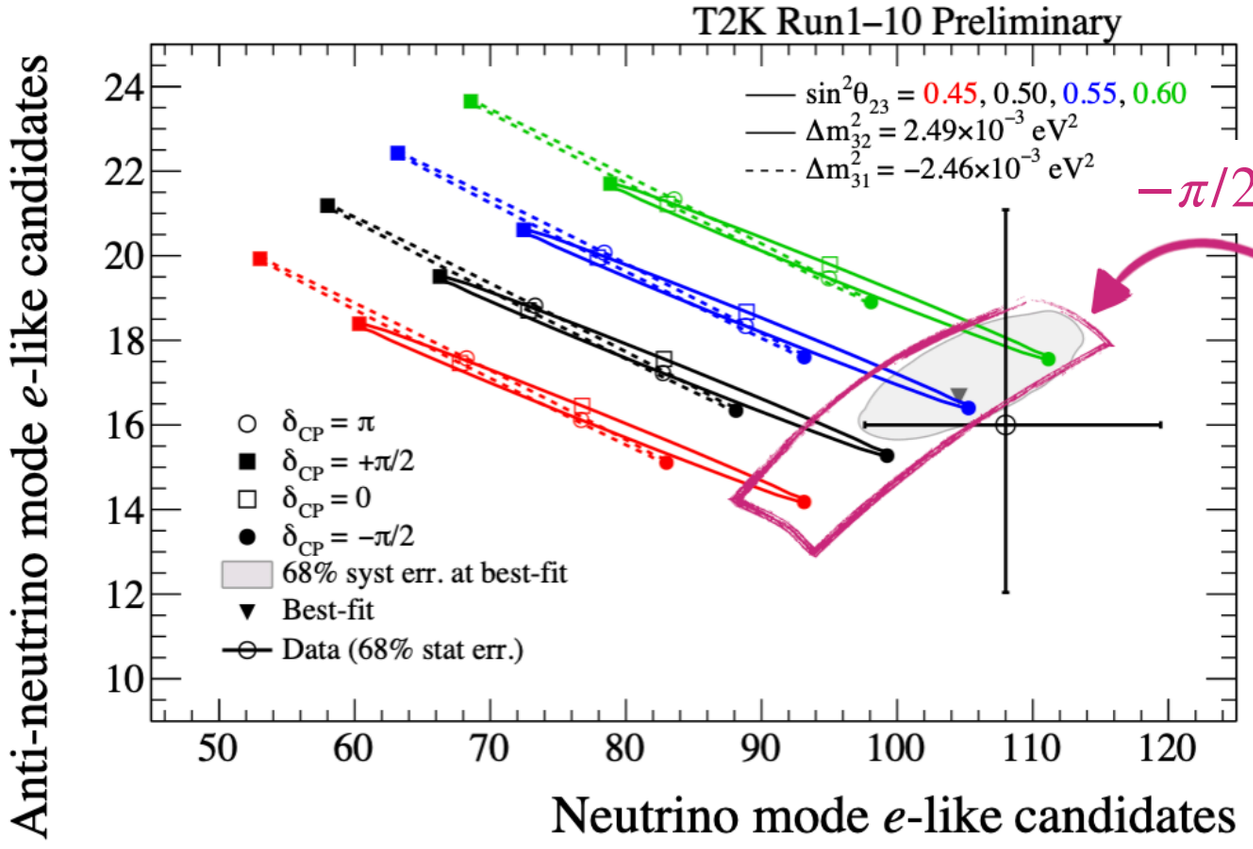
- Small amount of data → no 😭
  ν-mode: 119, $\bar{\nu}$-mode: 16



Can we use Wilk's theorem for leptonic CP violation search?

- Small amount of data → no 😭
  ν-mode: 119, $\bar{\nu}$-mode: 16

- Parameters with boundaries → no 😭
  e.g. $-1 \leq \sin\delta \leq 1$



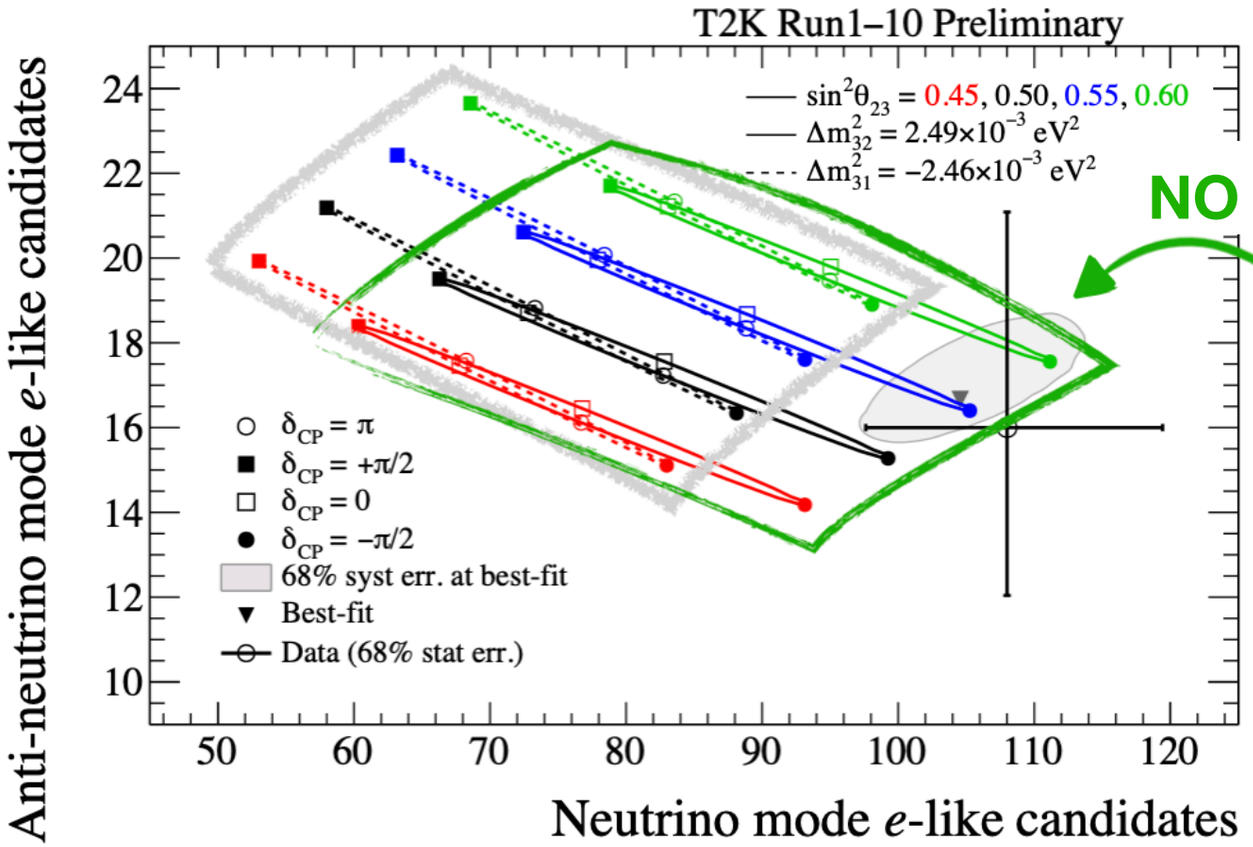
Can we use Wilk's theorem for leptonic CP violation search?

- Small amount of data → no 😭
  ν-mode: 119, $\bar{\nu}$-mode: 16

- Parameters with boundaries → no 😭
  e.g. $-1 \leq \sin\delta \leq 1$

- (approximate) degeneracies → no 😭
  e.g. mostly sensitive to $\sin\delta$
  = degenerate in $\pm\mathrm{sgn}\cos\delta$

  In fact an 8-fold degeneracy exists.



Can we use Wilk's theorem for leptonic CP violation search?

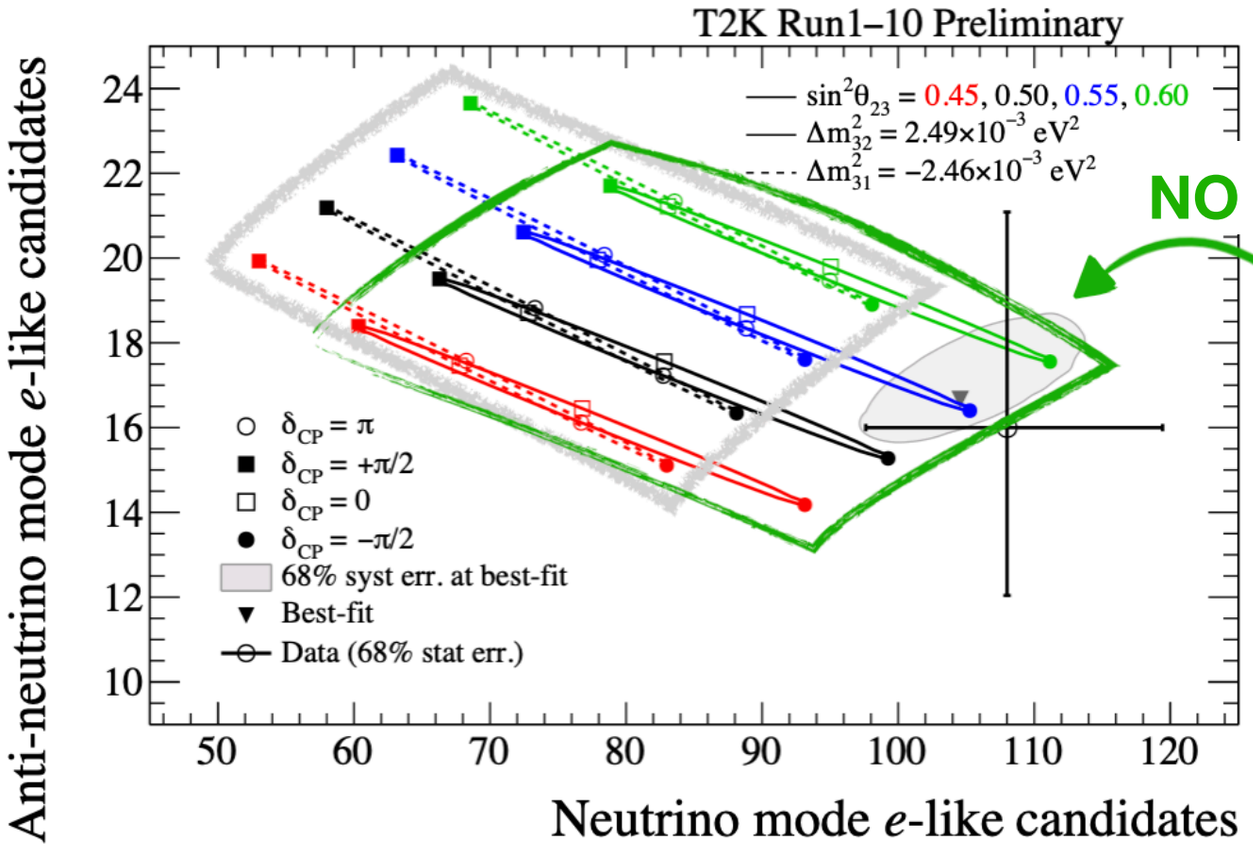
- Small amount of data → no 😭
  ν-mode: 119, $\bar{\nu}$-mode: 16

- Parameters with boundaries → no 😭
  e.g. $-1 \leq \sin\delta \leq 1$

- (approximate) degeneracies → no 😭
  e.g. mostly sensitive to $\sin\delta$
  = degenerate in $\pm\mathrm{sgn}\cos\delta$

  In fact an 8-fold degeneracy exists.

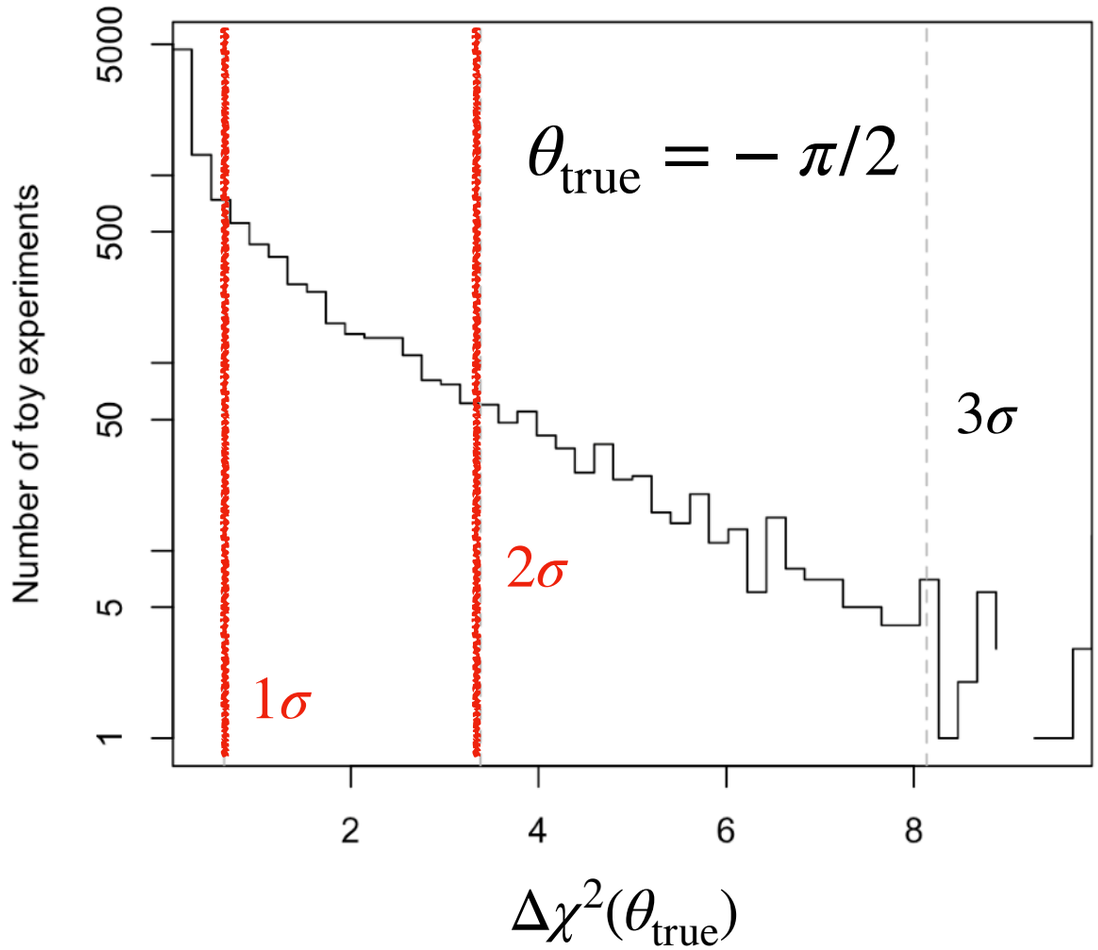
- θ contains (effectively) discrete parameters / hypotheses
  here: mass ordering → no 😭



Feldman-Cousins

Neyman-Construction using $\Delta\chi^2$ as ordering principle

- generate many toy experiments at each θ_{true}
- compute $\Delta\chi^2(\theta_{\text{true}})$ for all to obtain its distribution



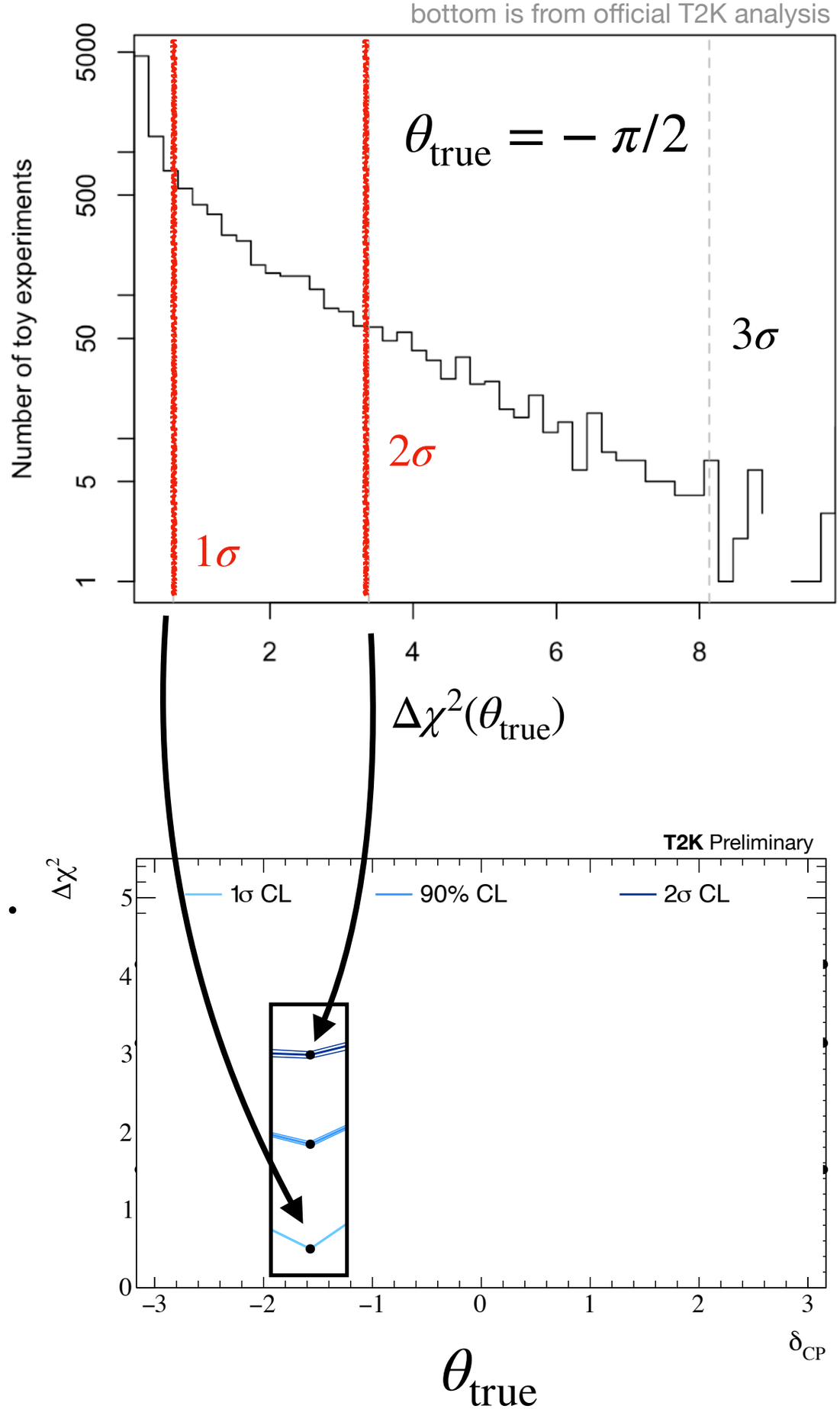
- get “critical values” s.t.
  $P\left(\Delta\chi^2_{\text{true}} < \Delta\chi^2_c \mid \theta_{\text{true}}\right) = 68\,\%, 90\,\%, \cdots$
  basically the percentiles as function of θ_{true}

Feldman-Cousins

Neyman-Construction using $\Delta\chi^2$ as ordering principle

- generate many toy experiments at each θ_{true}
- compute $\Delta\chi^2(\theta_{\text{true}})$ for all to obtain its distribution
- get “critical values” s.t. $P\left(\Delta\chi^2_{\text{true}} < \Delta\chi^2_c \mid \theta_{\text{true}}\right) = 68\,\%, 90\,\%, \cdots$ basically the percentiles as function of θ_{true}

Note: top plot is from a toy model, bottom is from official T2K analysis

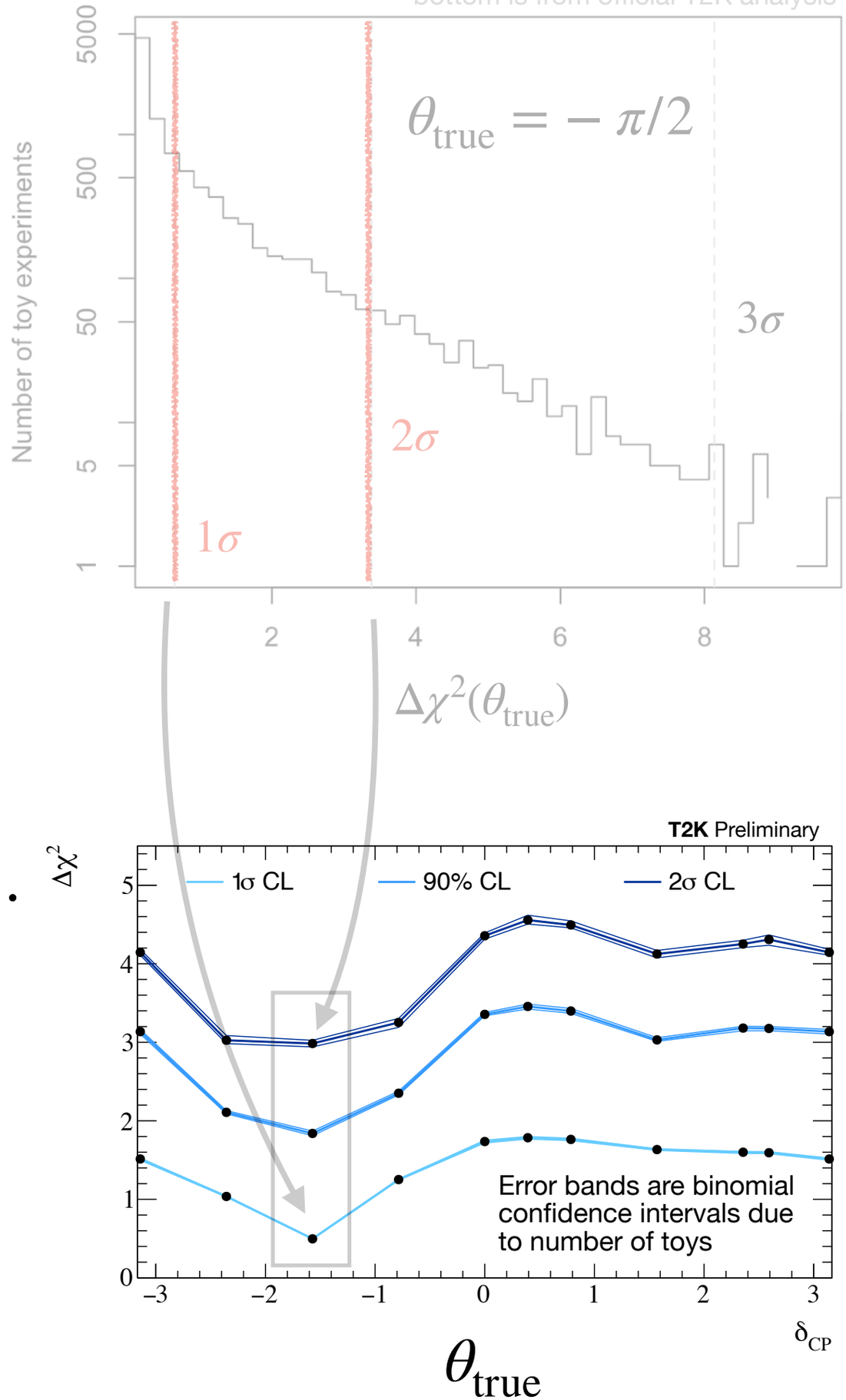


Feldman-Cousins

Neyman-Construction using $\Delta\chi^2$ as ordering principle

- generate many toy experiments at each θ_{true}
- compute $\Delta\chi^2(\theta_{\text{true}})$ for all to obtain its distribution
- get “critical values” s.t. $P\left(\Delta\chi^2_{\text{true}} < \Delta\chi^2_c \mid \theta_{\text{true}}\right) = 68\,\%, 90\,\%, \dots$ basically the percentiles as function of θ_{true}

Note: top plot is from a toy model, bottom is from official T2K analysis

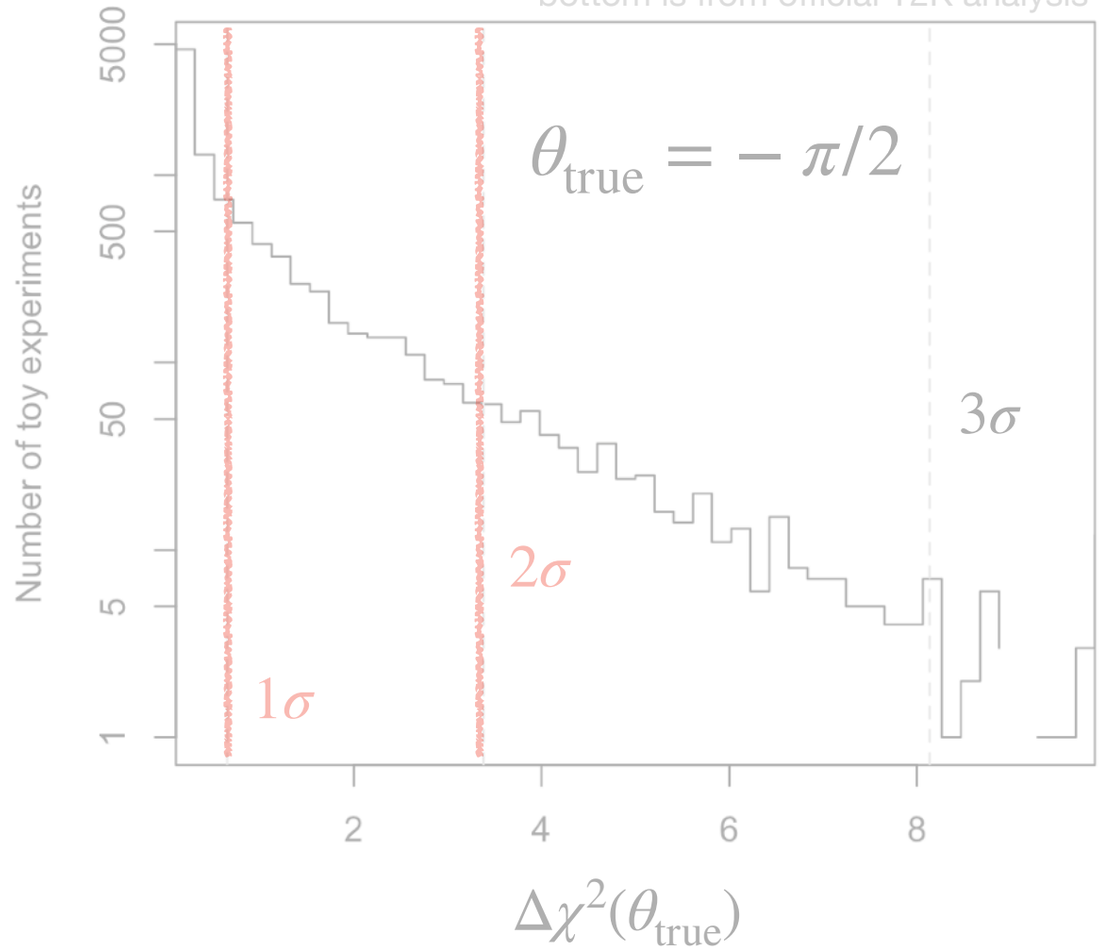


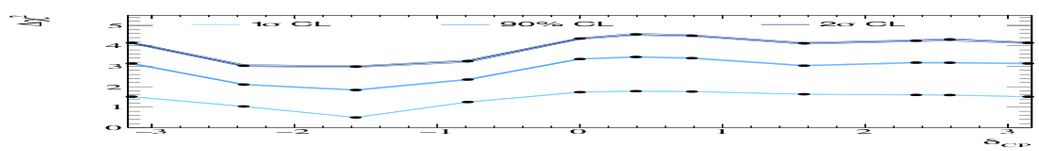
Feldman-Cousins

Neyman-Construction using $\Delta\chi^2$ as ordering principle

- generate many toy experiments at each θ_{true}
- compute $\Delta\chi^2(\theta_{\text{true}})$ for all to obtain its distribution
- get “critical values” s.t.
  $P\left(\Delta\chi^2_{\text{true}} < \Delta\chi^2_c \mid \theta_{\text{true}}\right) = 68\,\%, 90\,\%, \cdots$
  basically the percentiles as function of θ_{true}

Note: top plot is from a toy model, bottom is from official T2K analysis





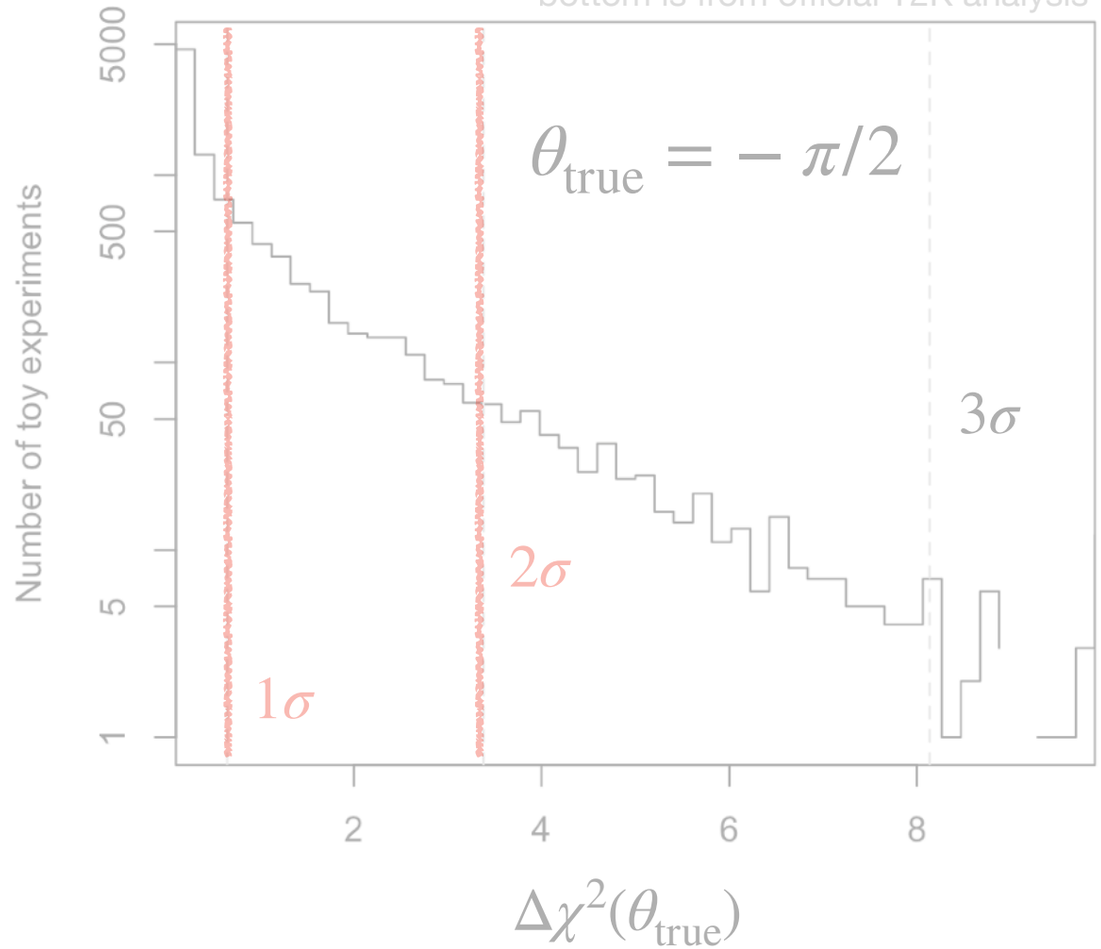
T2K Preliminary

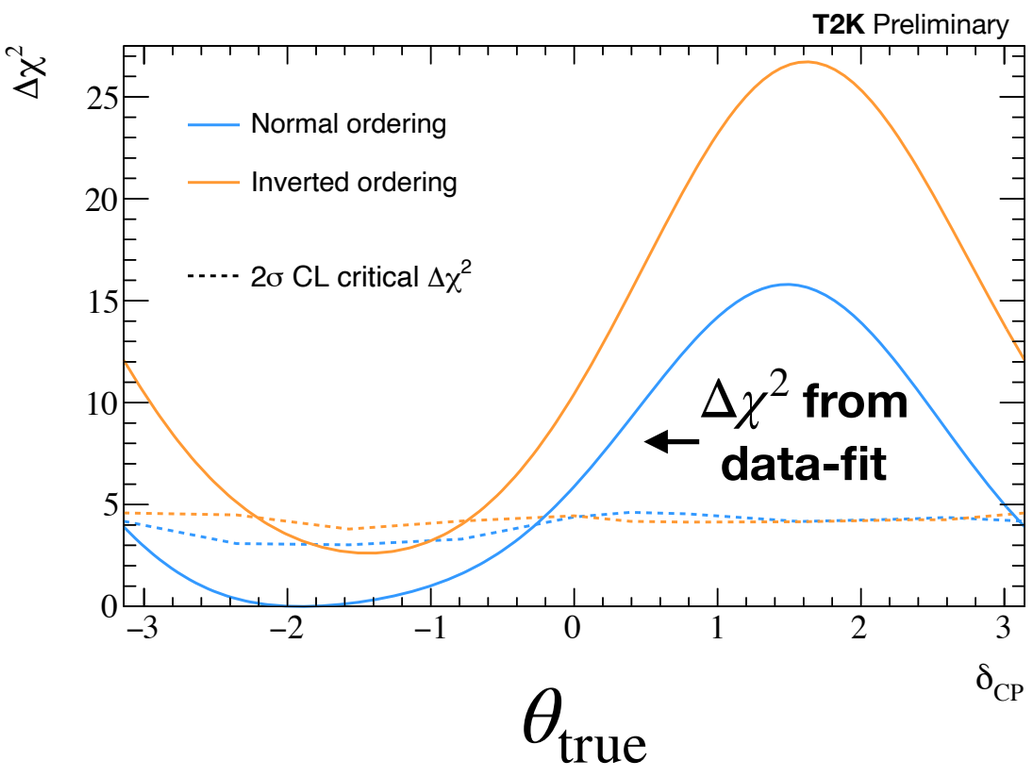
Feldman-Cousins

Neyman-Construction using $\Delta\chi^2$ as ordering principle

- generate many toy experiments at each θ_{true}
- compute $\Delta\chi^2(\theta_{\text{true}})$ for all to obtain its distribution
- get “critical values” s.t. $P\left(\Delta\chi^2_{\text{true}} < \Delta\chi^2_c \mid \theta_{\text{true}}\right) = 68\,\%, 90\,\%, \cdots$ basically the percentiles as function of θ_{true}
- construct confidence interval by $\Delta\chi^2(\theta) < \Delta\chi^2_c(\theta)$

Note: top plot is from a toy model, bottom is from official T2K analysis



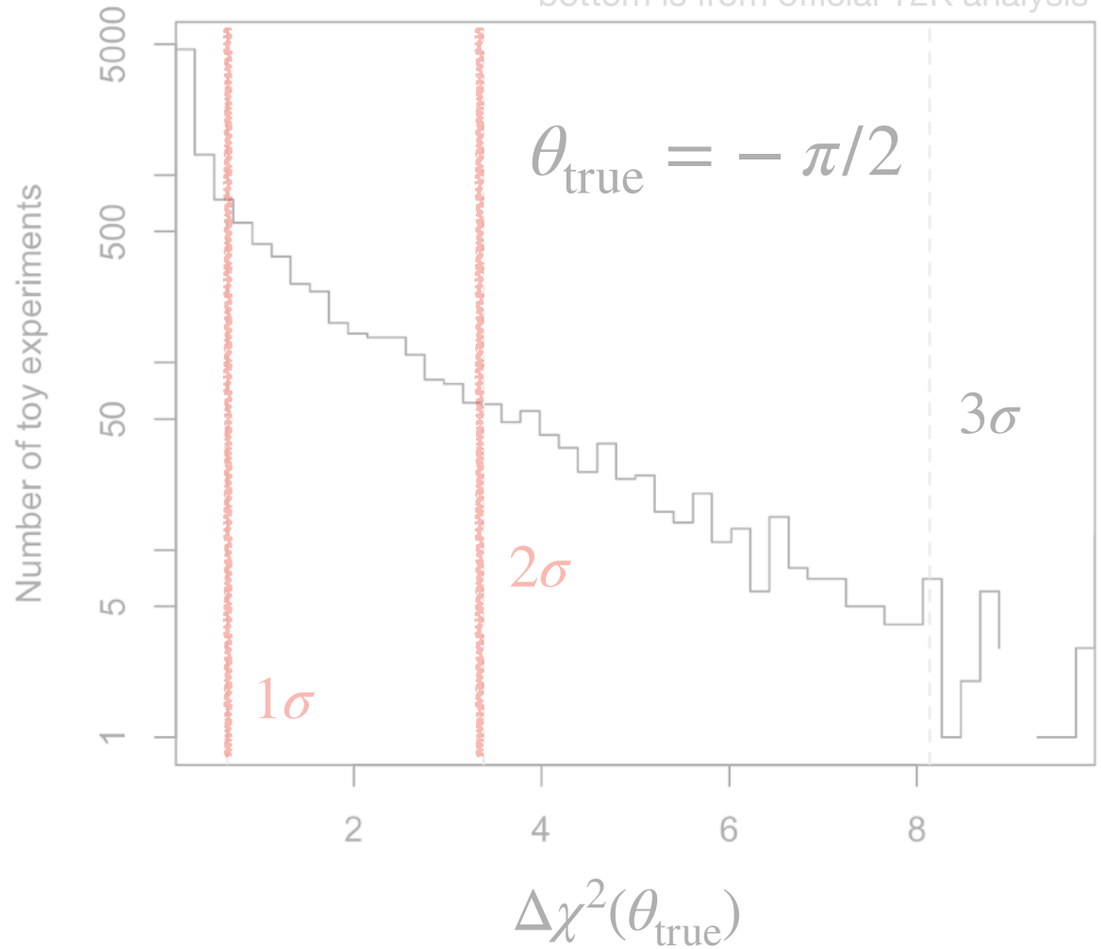


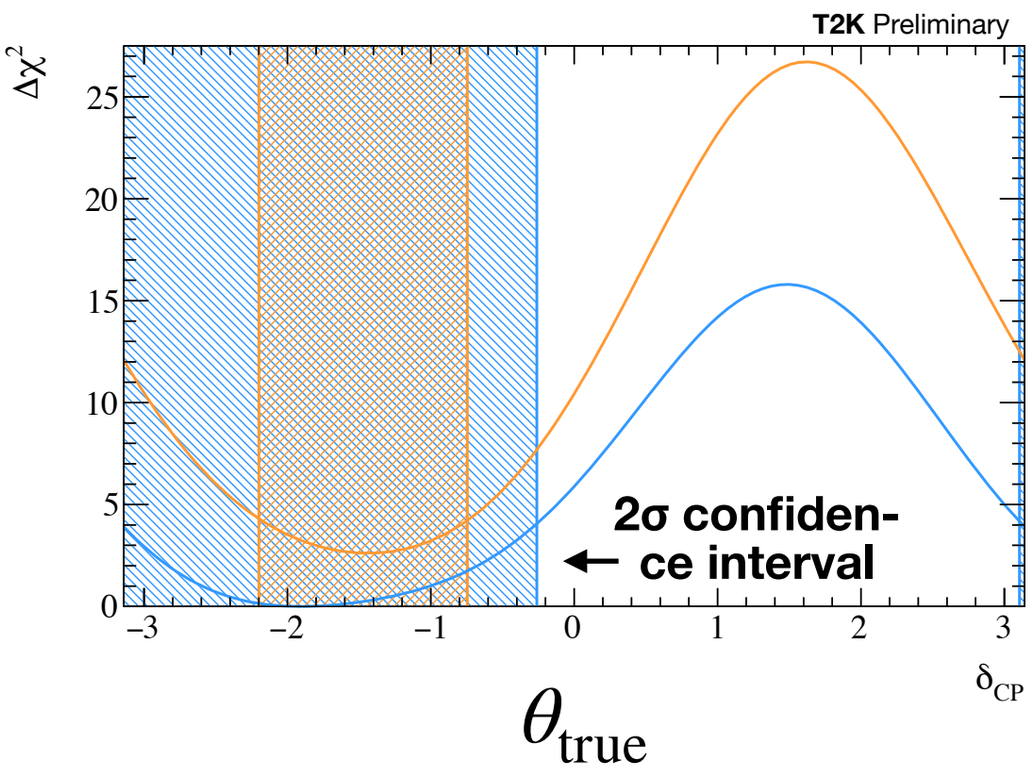
Feldman-Cousins

Neyman-Construction using $\Delta\chi^2$ as ordering principle

- generate many toy experiments at each θ_{true}
- compute $\Delta\chi^2(\theta_{\text{true}})$ for all to obtain its distribution
- get “critical values” s.t. $P\left(\Delta\chi^2_{\text{true}} < \Delta\chi^2_c \mid \theta_{\text{true}}\right) = 68\,\%, 90\,\%, \cdots$ basically the percentiles as function of θ_{true}
- construct confidence interval by $\Delta\chi^2(\theta) < \Delta\chi^2_c(\theta)$

Note: top plot is from a toy model, bottom is from official T2K analysis

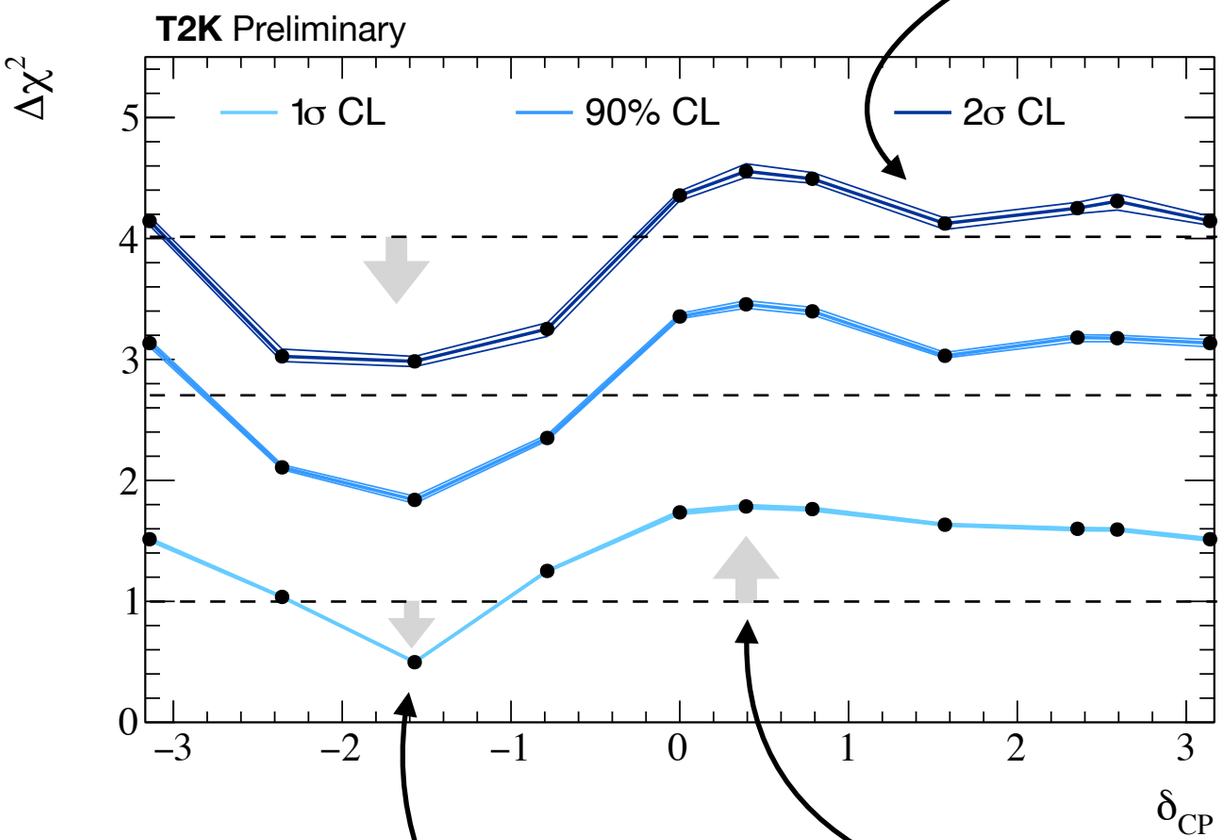




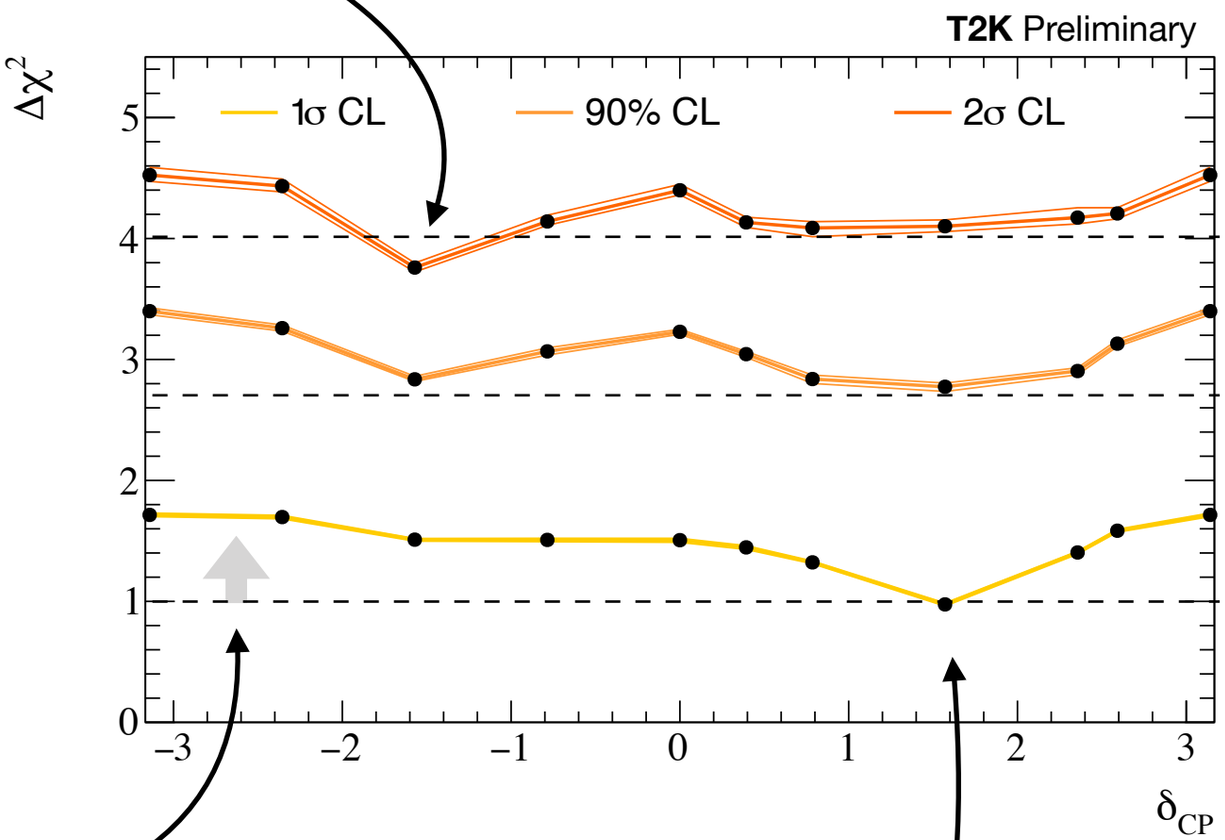
Critical values for δ_{CP}

At high CL start to see the $\sin\delta$ boundary in other mass ordering

Normal ordering



Inverted ordering



MO and $\mathrm{sgn}\cos\delta$ degeneracies

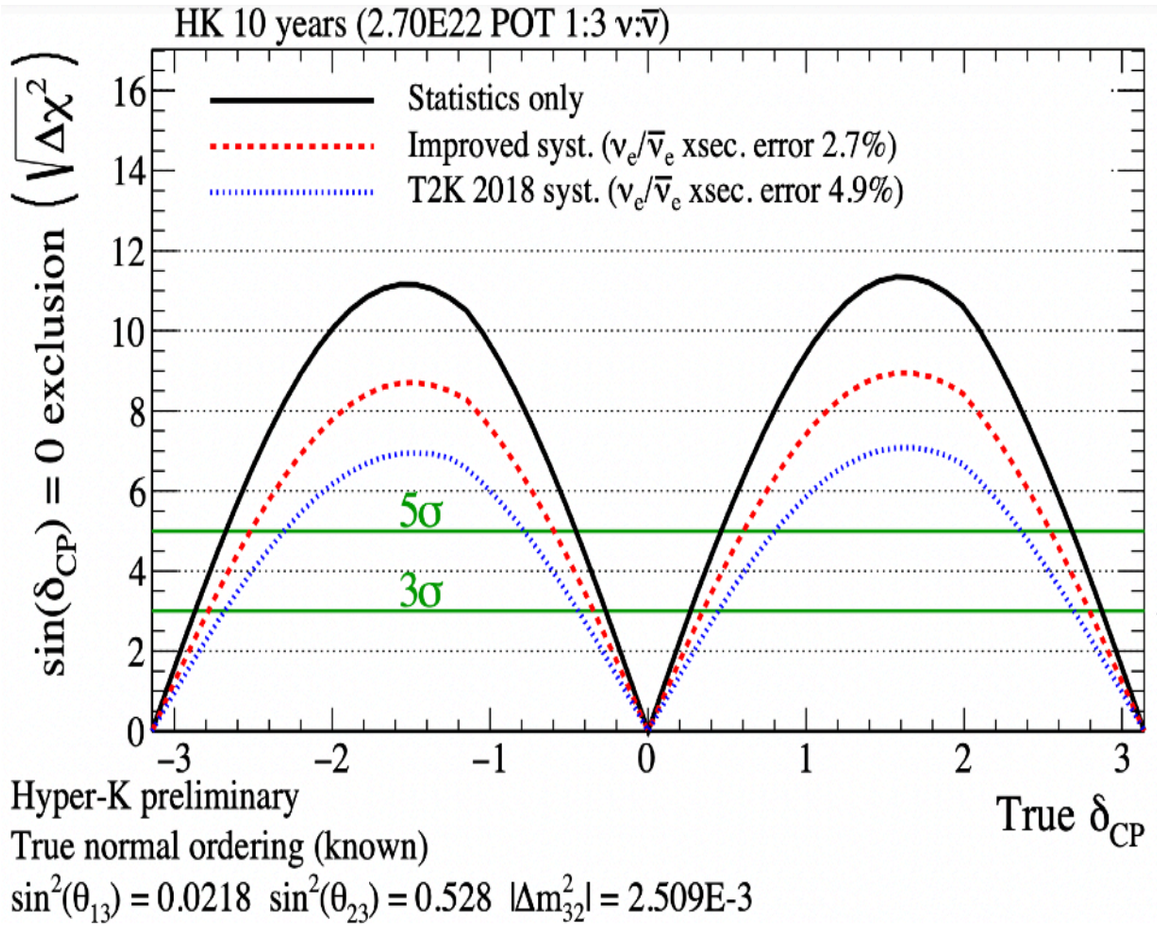
$\sin\delta = -1$ boundary

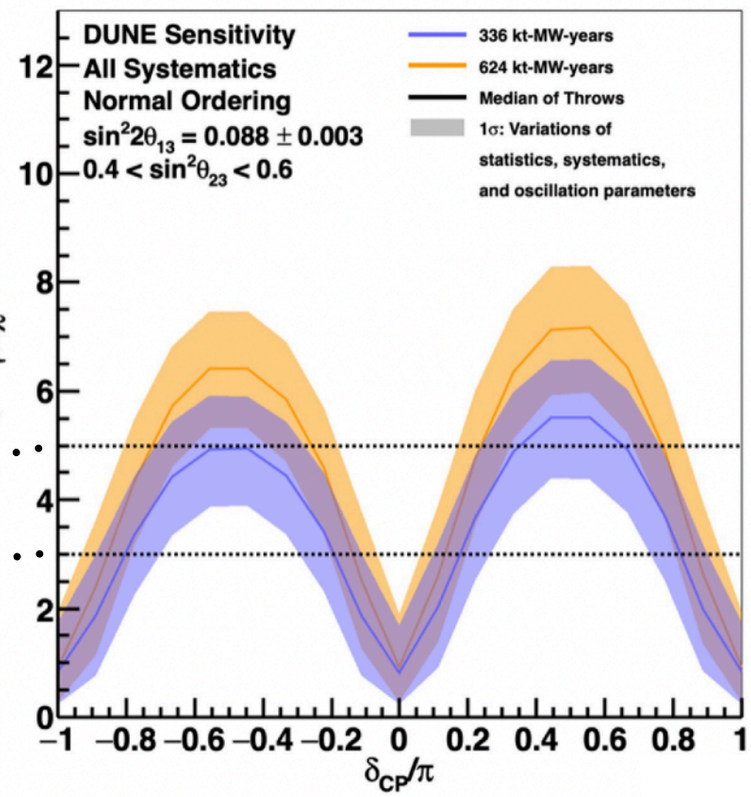
$\sin\delta = +1$ boundary

from L. Berns, Moriond 2021

Problem

- Current experiments (T2K) started excluding δ_{CP} ranges at 3σ CL.
  Future experiments (HK, DUNE) aiming for 5σ discovery of CP violation in leptons.





J. Wilson, Neutrino 2022

M. Muether, Neutrino 2022

Problem

- Current experiments (T2K) started excluding δ_{CP} ranges at 3σ CL.
  Future experiments (HK, DUNE) aiming for 5σ discovery of CP violation in leptons.

- By definition, to obtain high confidence-level critical values in Feldman-Cousins method, need many toy experiments:

$3\sigma : \gg 370$
$4\sigma : \gg 16\mathrm{k}$
$5\sigma : \gg 1.7\mathrm{M}$

each fit is complicated with $\mathcal{O}(100)$ parameters often having non-linear responses, so this is not very practical.

Problem

- Current experiments (T2K) started excluding δ_{CP} ranges at 3σ CL.
  Future experiments (HK, DUNE) aiming for 5σ discovery of CP violation in leptons.

- By definition, to obtain high confidence-level critical values in Feldman-Cousins method, need many toy experiments:

  $3\sigma : \gg 370$
  $4\sigma : \gg 16\mathrm{k}$
  $5\sigma : \gg 1.7\mathrm{M}$

  each fit is complicated with $\mathcal{O}(100)$ parameters often having non-linear responses, so this is not very practical.

- Now even if we produce say 100M toy experiments to get 5σ critical values, most of the toy experiments will be around $1\sigma \sim 2\sigma$ and “wasted”.

- Can we do better? Can we specifically generate toys that matter at high CL?

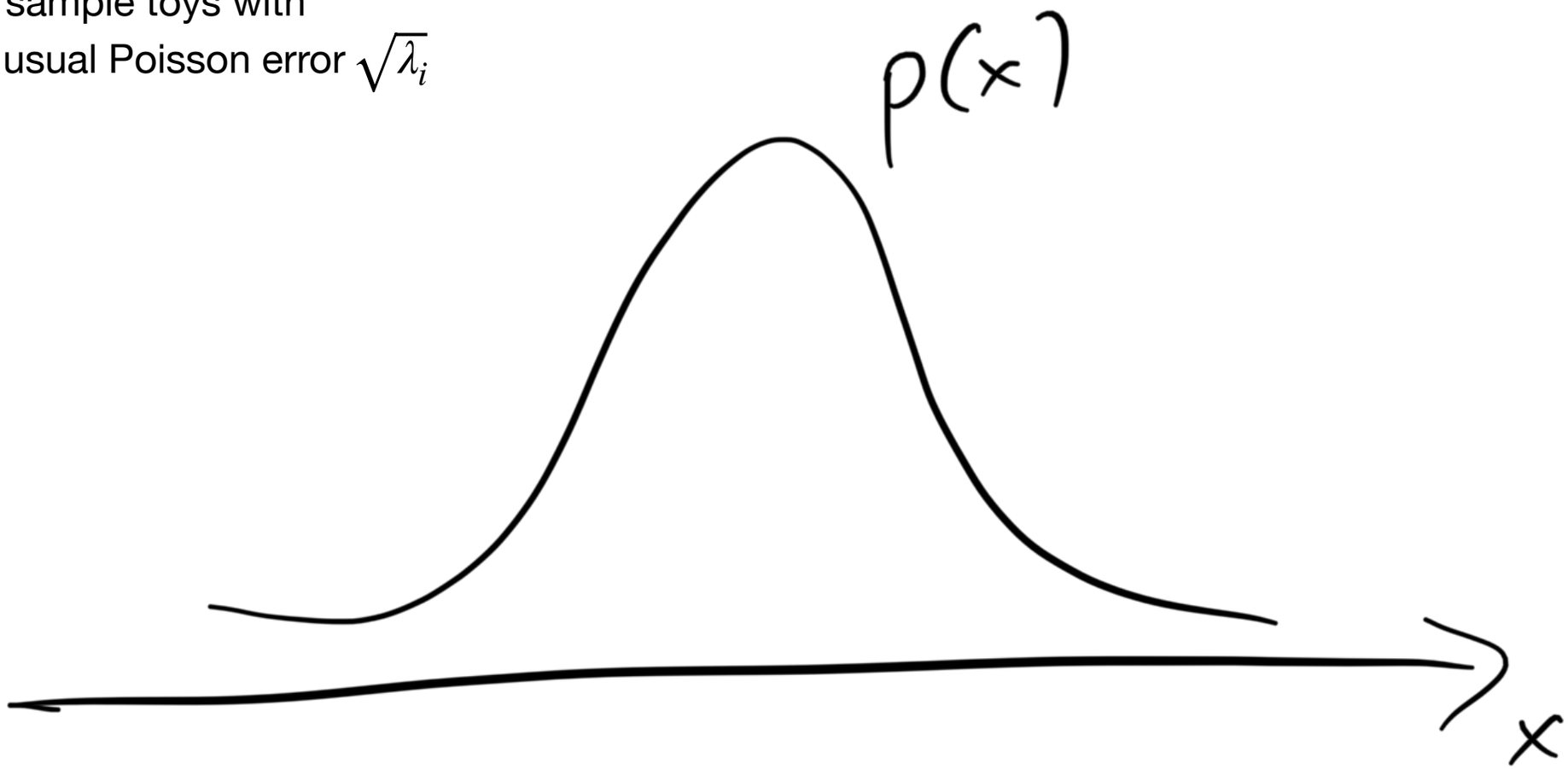
Naive attempt

- Let's preferentially sample all low-probability toys "increase the temperature T"

  e.g. $p_{\text{smpl}}(x) := [p(x \mid \theta_{\text{true}})]^{1/T}$

  and weight each toy as $w(x) = \dfrac{p(x \mid \theta_{\text{true}})}{p_{\text{smpl}}(x)}$

- So e.g. for each bin i, sample toys with $T\sqrt{\lambda_i}$ error instead of usual Poisson error $\sqrt{\lambda_i}$



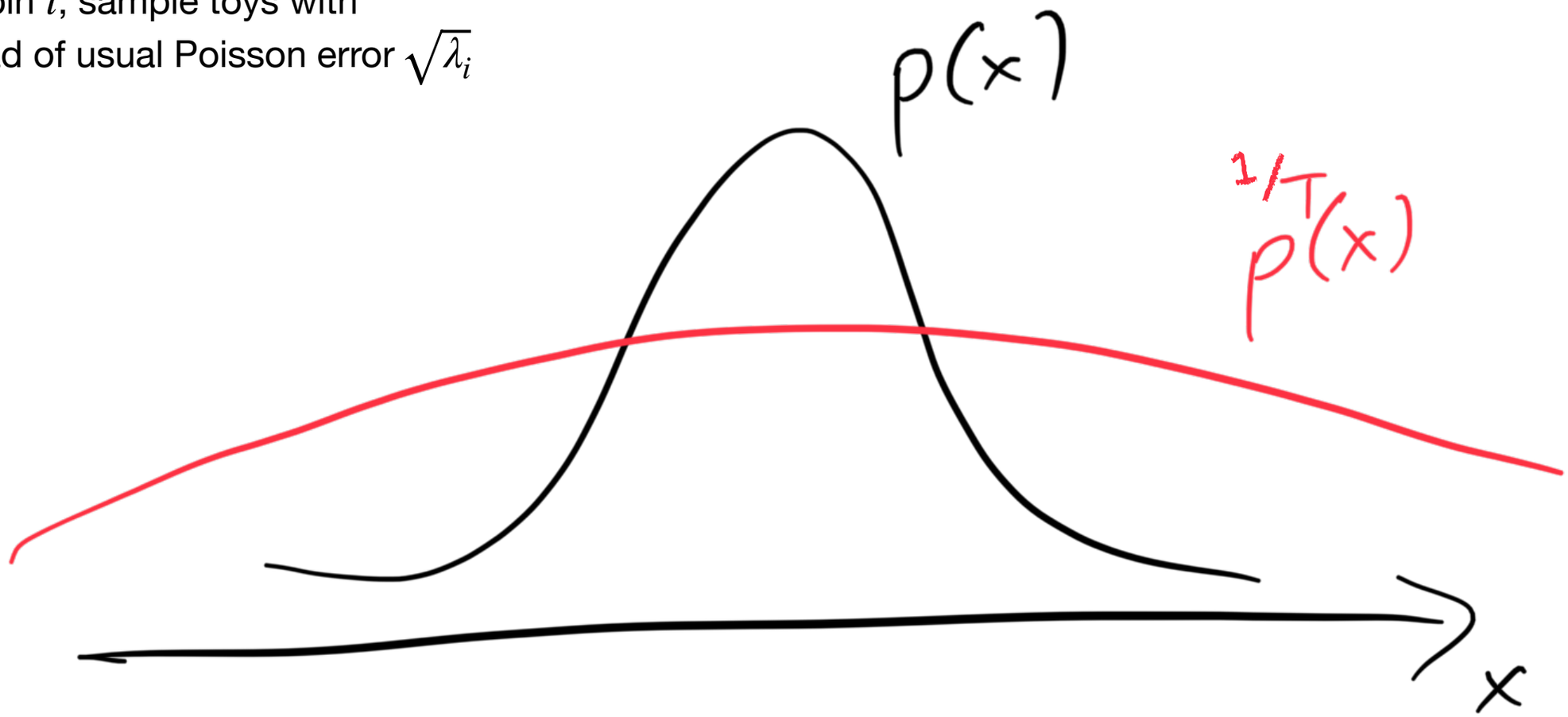
Naive attempt

- Let's preferentially sample all low-probability toys "increase the temperature T"

  e.g. $p_{\text{smpl}}(x) := [p(x \mid \theta_{\text{true}})]^{1/T}$

  and weight each toy as $w(x) = \dfrac{p(x \mid \theta_{\text{true}})}{p_{\text{smpl}}(x)}$

- So e.g. for each bin i, sample toys with $T\sqrt{\lambda_i}$ error instead of usual Poisson error $\sqrt{\lambda_i}$



Naive attempt

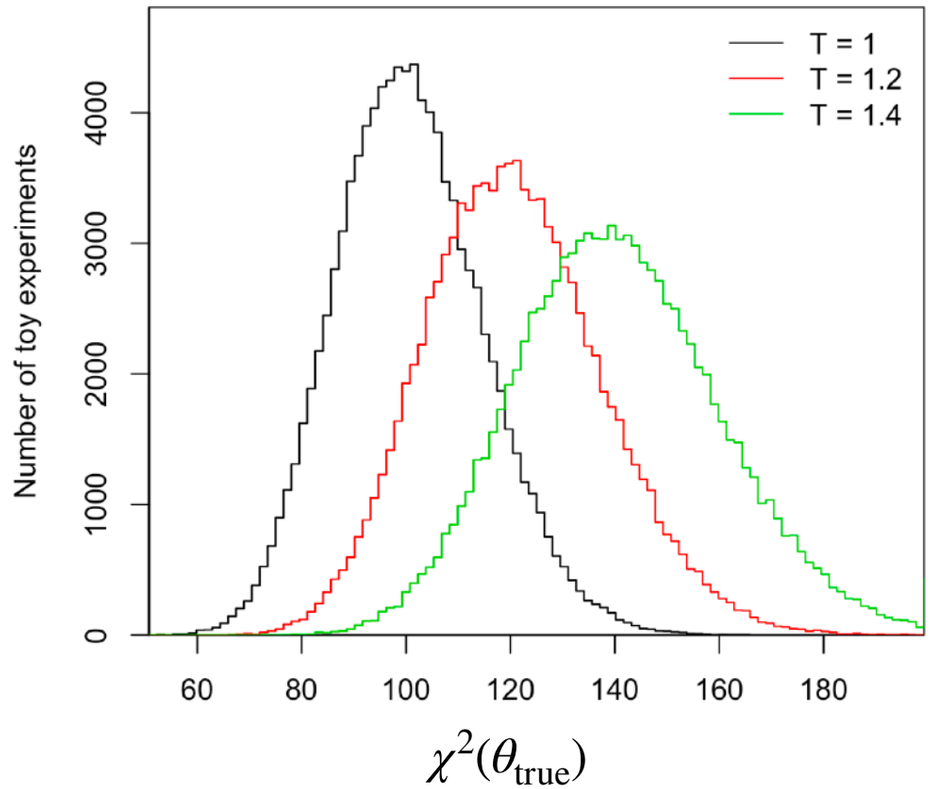
- Let's preferentially sample all low-probability toys "increase the temperature T"

  e.g. $p_{\text{smpl}}(x) := [p(x \mid \theta_{\text{true}})]^{1/T}$

  and weight each toy as $w(x) = \dfrac{p(x \mid \theta_{\text{true}})}{p_{\text{smpl}}(x)}$

- So e.g. for each bin i, sample toys with $T\sqrt{\lambda_i}$ error instead of usual Poisson error $\sqrt{\lambda_i}$

- This gives you toys with larger values of χ^2



Naive attempt

- Let's preferentially sample all low-probability toys "increase the temperature T"

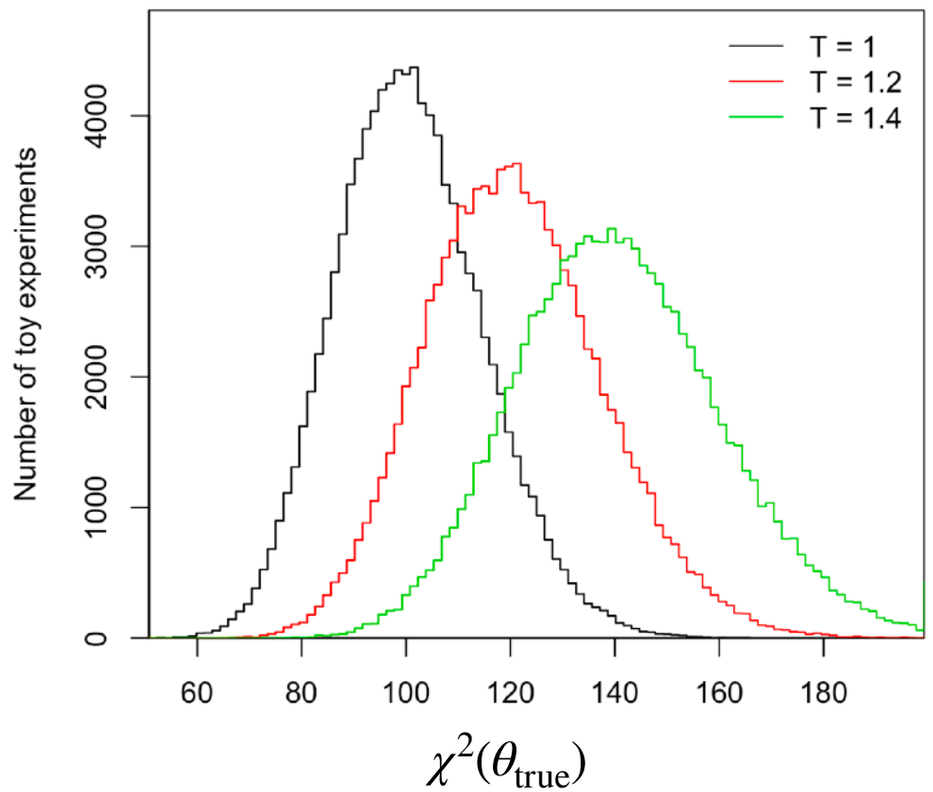
  e.g. $p_{\text{smpl}}(x) := [p(x \mid \theta_{\text{true}})]^{1/T}$

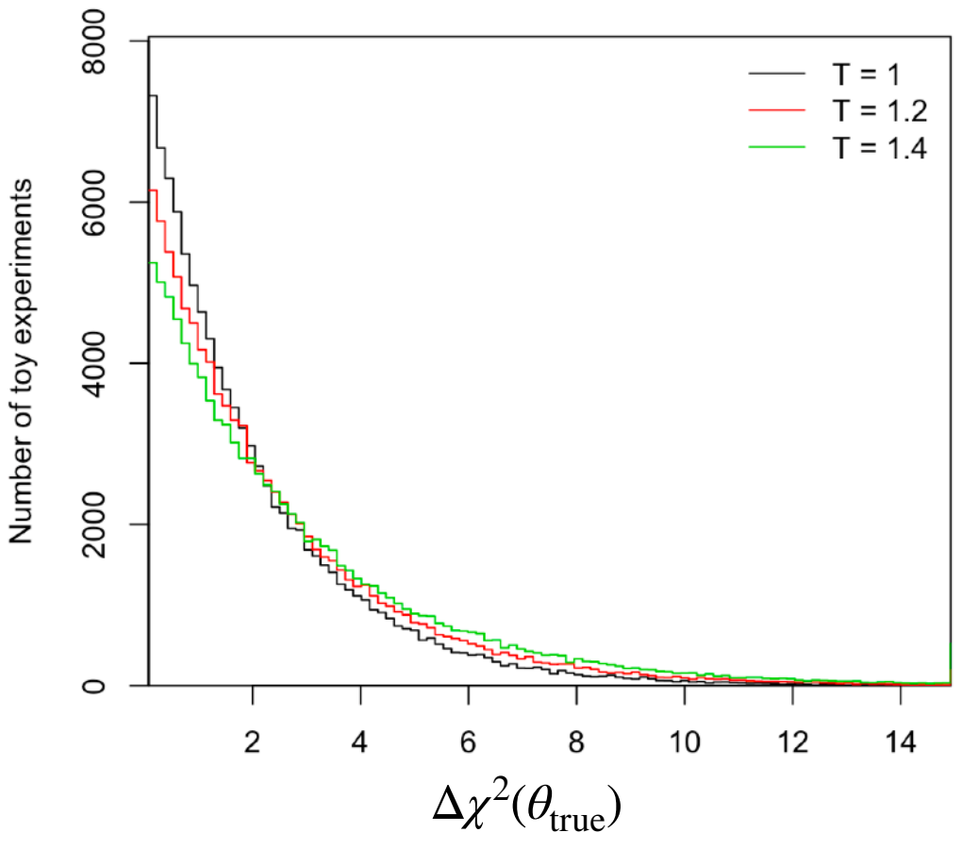
  and weight each toy as $w(x) = \dfrac{p(x \mid \theta_{\text{true}})}{p_{\text{smpl}}(x)}$

- So e.g. for each bin i, sample toys with $T\sqrt{\lambda_i}$ error instead of usual Poisson error $\sqrt{\lambda_i}$

- This gives you toys with larger values of χ^2

- But! almost all of them have small $\Delta\chi^2(\theta_{\text{true}})$, because there are N_{bins} directions in which we can increase the temperature, but only k (non-linear) directions of them contribute to a change of $\Delta\chi^2(\theta)$, the rest will only increase χ^2_{min}





Naive attempt

- Let's preferentially sample all low-probability toys "increase the temperature T"

  e.g. $p_{\text{smpl}}(x) := [p(x \mid \theta_{\text{true}})]^{1/T}$

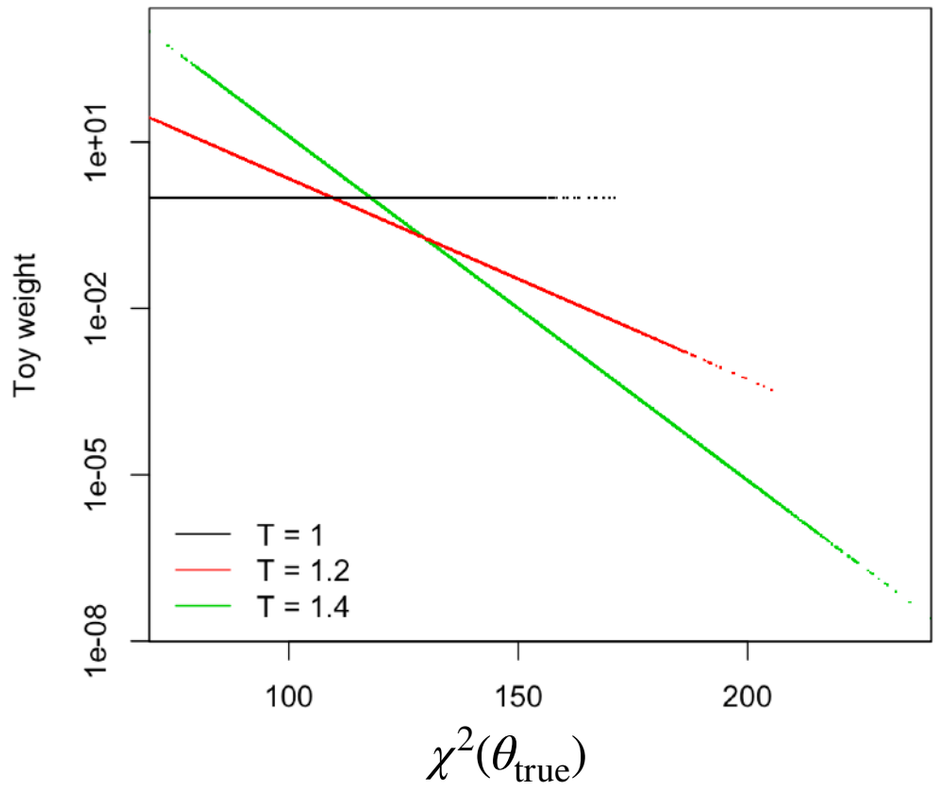
  and weight each toy as $w(x) = \dfrac{p(x \mid \theta_{\text{true}})}{p_{\text{smpl}}(x)}$

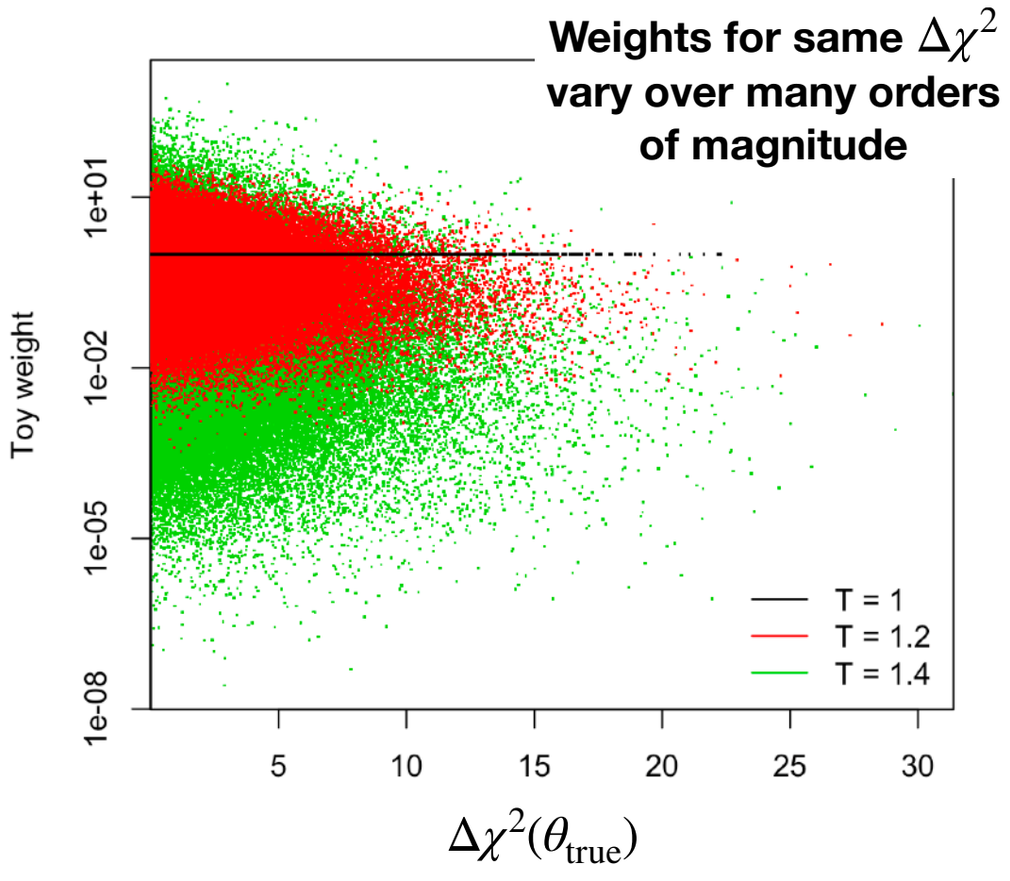
- So e.g. for each bin i, sample toys with $T\sqrt{\lambda_i}$ error instead of usual Poisson error $\sqrt{\lambda_i}$

- This gives you toys with larger values of χ^2

- But! almost all of them have small $\Delta\chi^2(\theta_{\text{true}})$, because there are N_{bins} directions in which we can increase the temperature, but only k (non-linear) directions of them contribute to a change of $\Delta\chi^2(\theta)$, the rest will only increase χ^2_{min}

- In fact, because now we have to weight each toy x with $w(x)$, any variation in $w(x)$ for the same value of $\Delta\chi^2$ will actually "cost" us toy statistics, so this is even worse than before, not only at low $\Delta\chi^2$, but also at high $\Delta\chi^2$





Naive attempt

- Let's preferentially sample all low-probability toys "increase the temperature T"

  e.g. $p_{\text{smpl}}(x) := [p(x \mid \theta_{\text{true}})]^{1/T}$

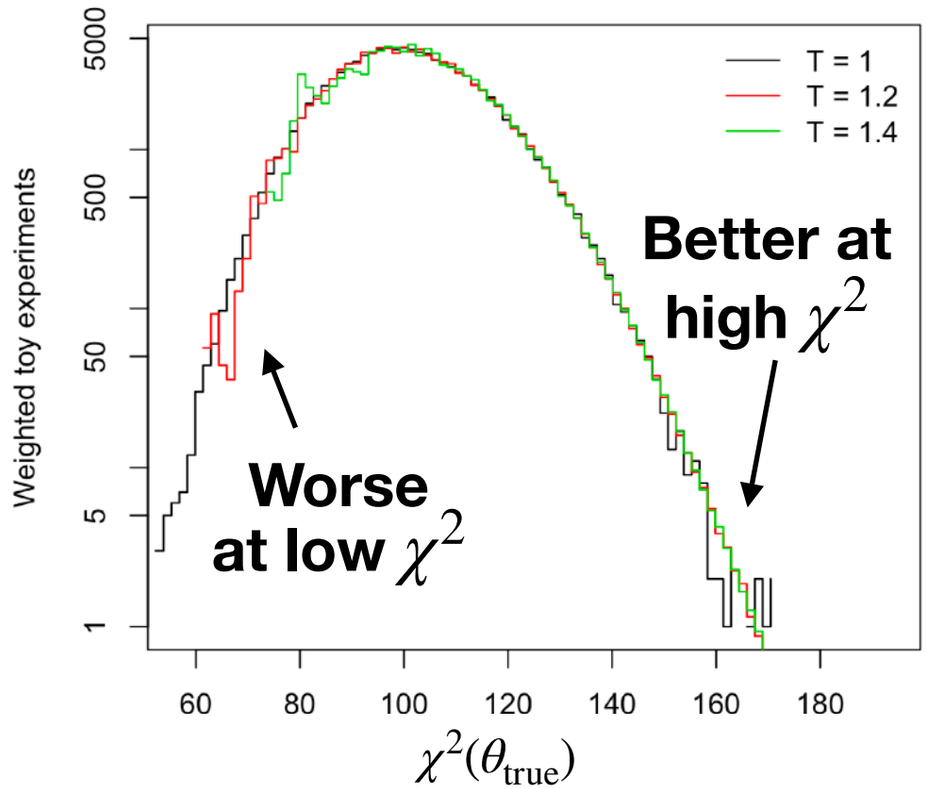
  and weight each toy as $w(x) = \dfrac{p(x \mid \theta_{\text{true}})}{p_{\text{smpl}}(x)}$

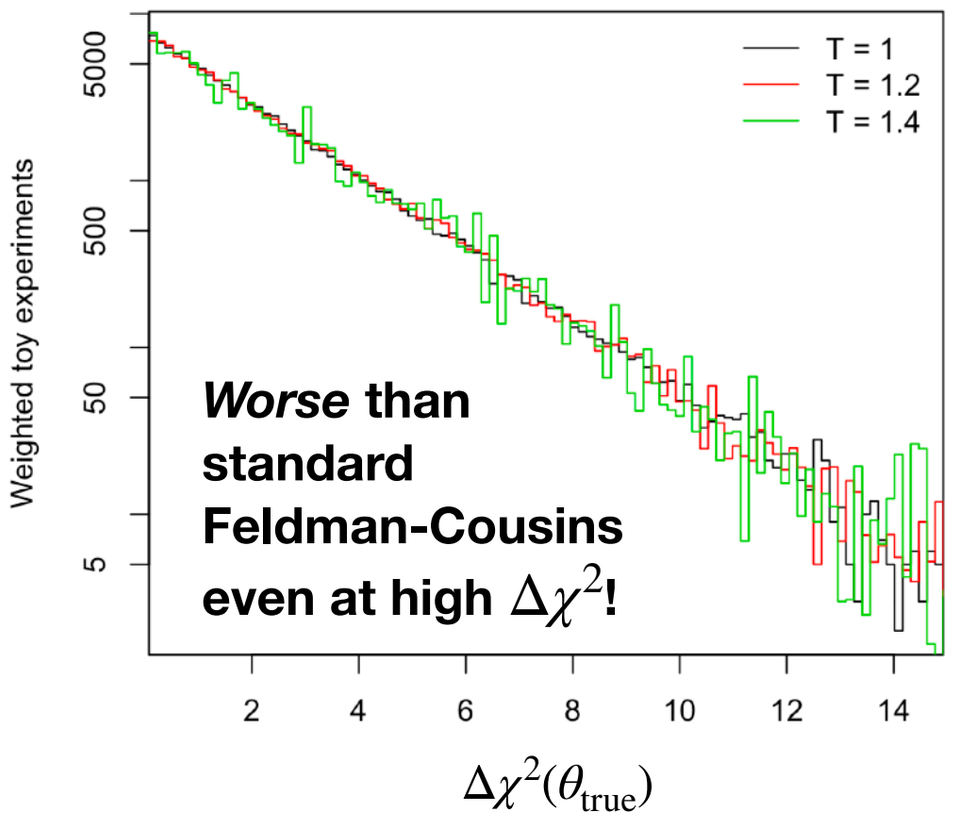
- So e.g. for each bin i, sample toys with $T\sqrt{\lambda_i}$ error instead of usual Poisson error $\sqrt{\lambda_i}$

- This gives you toys with larger values of χ^2

- But! almost all of them have small $\Delta\chi^2(\theta_{\text{true}})$, because there are N_{bins} directions in which we can increase the temperature, but only k (non-linear) directions of them contribute to a change of $\Delta\chi^2(\theta)$, the rest will only increase χ^2_{min}

- In fact, because now we have to weight each toy x with $w(x)$, any variation in $w(x)$ for the same value of $\Delta\chi^2$ will actually "cost" us toy statistics, so this is even worse than before, not only at low $\Delta\chi^2$, but also at high $\Delta\chi^2$



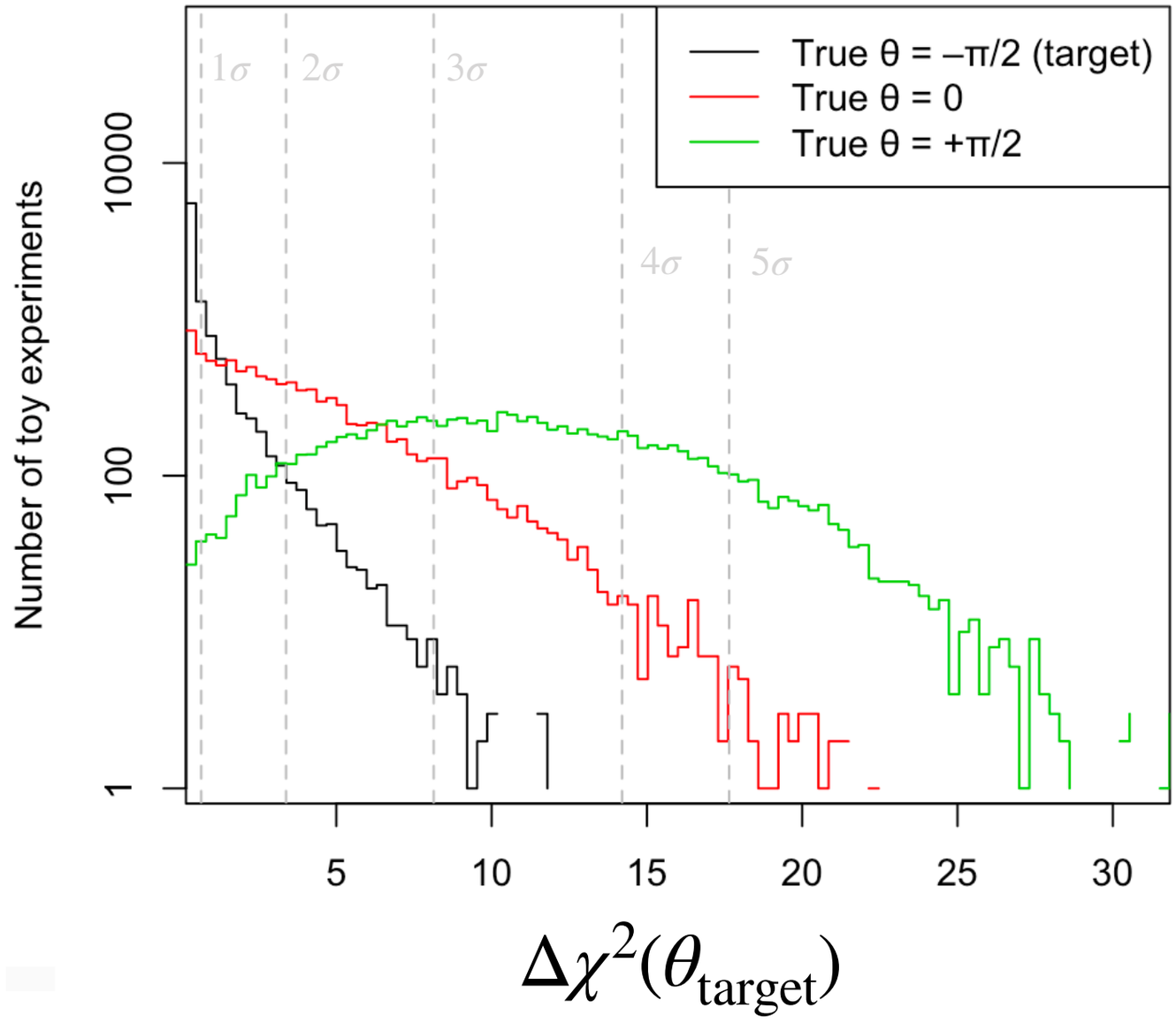


The solution: mixture of many θ

- What does high $\Delta\chi^2(\theta_{\text{true}})$ mean?
  → there exists a different value $\hat{\theta}$ from which one is more likely to obtain the data x than from θ_{true}

The solution: mixture of many θ

- What does high $\Delta\chi^2(\theta_{\text{true}})$ mean?
  → there exists a different value $\hat{\theta}$ from which one is more likely to obtain the data x than from θ_{true}



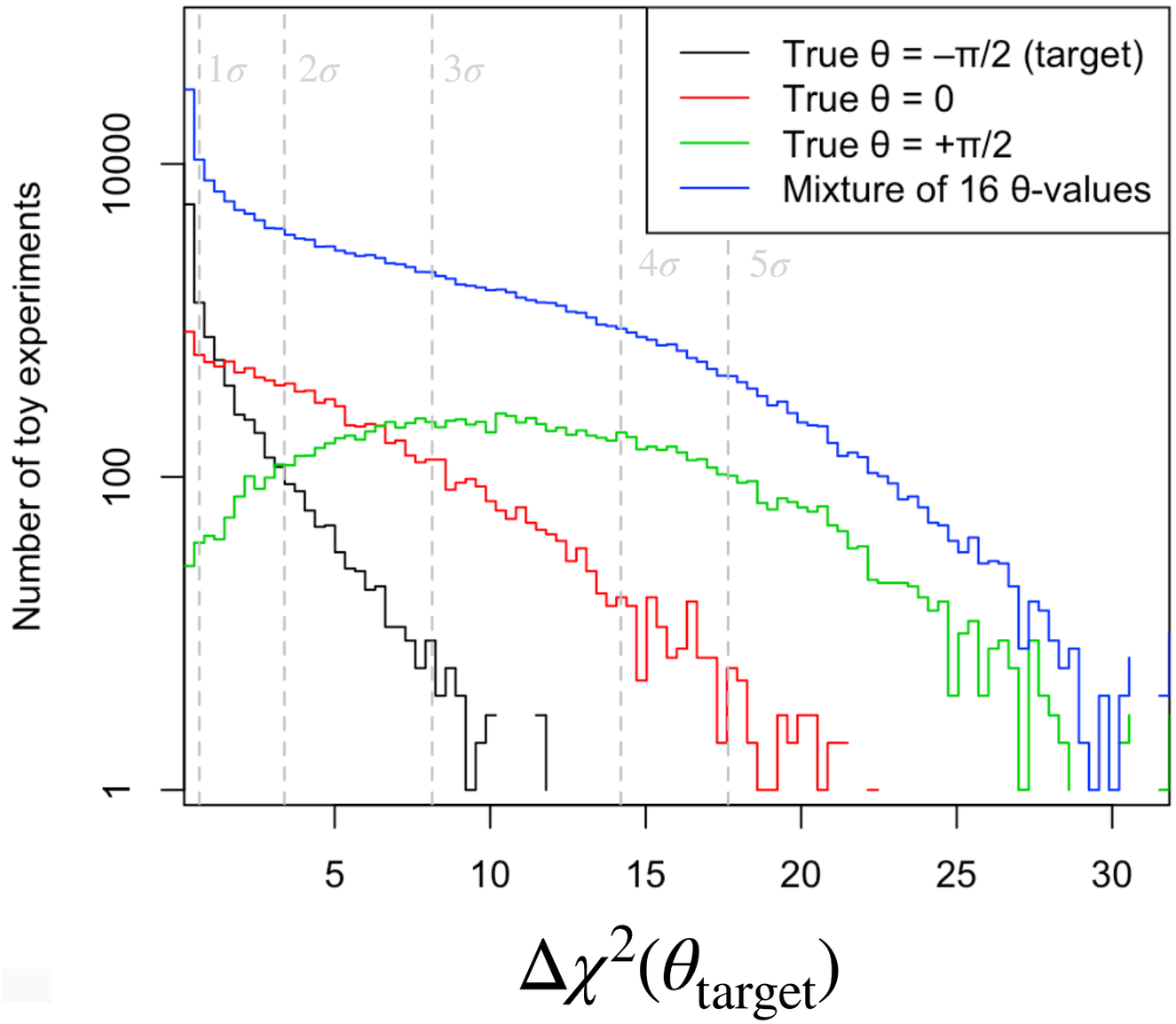
The solution: mixture of many θ

- What does high $\Delta\chi^2(\theta_{\text{true}})$ mean?
  → there exists a different value $\hat{\theta}$ from which one is more likely to obtain the data x than from θ_{true}

- Just use all toys sampled from many θ_{smpl} at once

$$p_{\text{smpl}}(x) = p(x \mid \{\theta\}) := \frac{1}{S}\sum_{s=1}^{S} p(x \mid \theta_s)$$

(mixture model)



The solution: mixture of many θ

- What does high $\Delta\chi^2(\theta_{\text{true}})$ mean?
  → there exists a different value $\hat{\theta}$ from which one is more likely to obtain the data x than from θ_{true}

- Just use all toys sampled from many θ_{smpl} at once

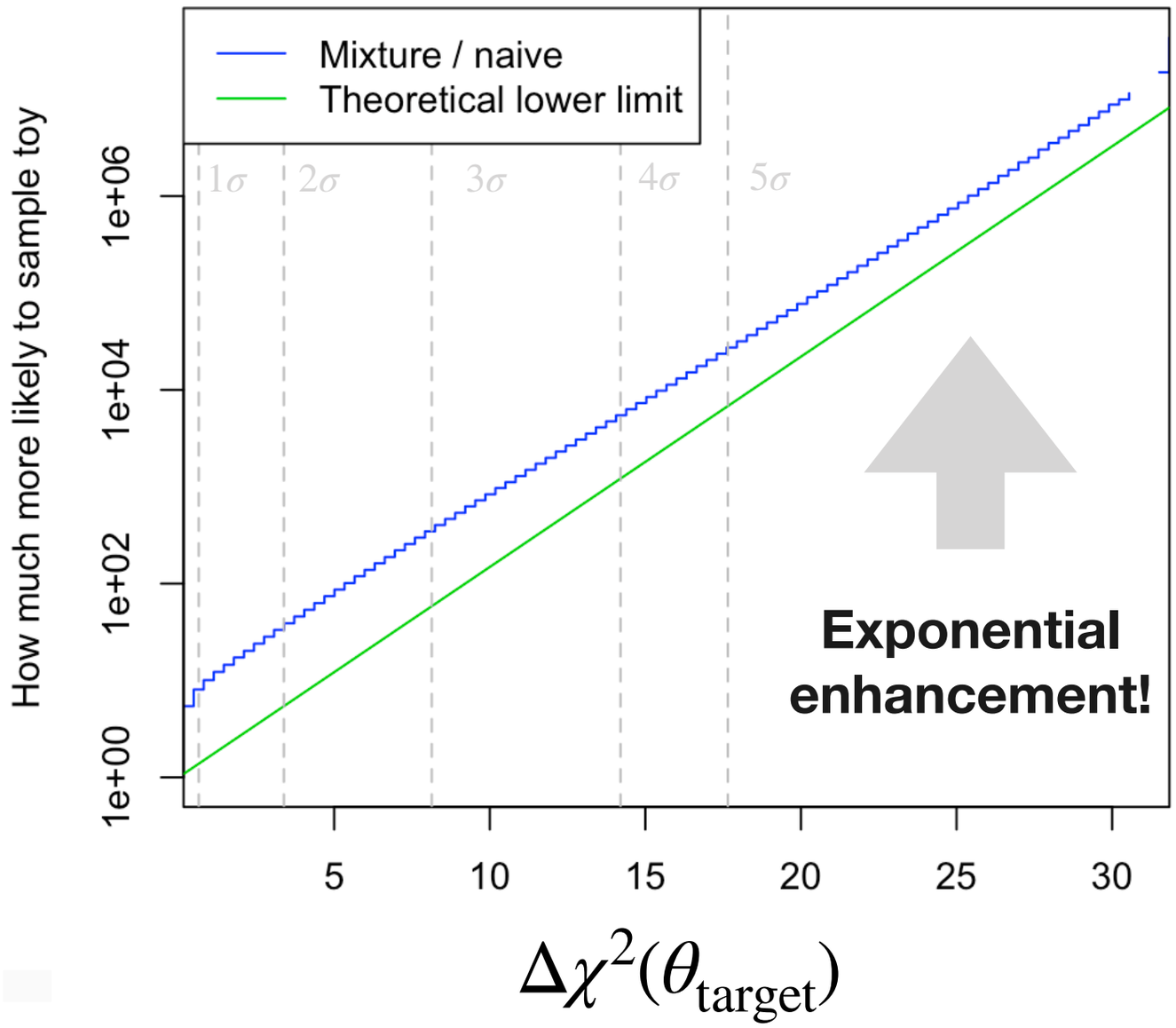
$$p_{\text{smpl}}(x) = p(x \mid \{\theta\}) := \frac{1}{S}\sum_{s=1}^{S} p(x \mid \theta_s)$$

(mixture model)

As long as θ_s are close enough, there will always be some θ_s close enough to $\hat{\theta}$

$$\forall x,\ \Delta\chi^2(\theta_{\text{target}}) \geq \Delta\chi^2_0$$

$$\frac{\mathrm{d}N_{\text{toys}}(x \mid \{\theta\})}{\mathrm{d}N_{\text{toys}}(x \mid \theta_{\text{target}})} = \frac{\sum_s p(x \mid \theta_s)}{p(x \mid \theta_{\text{target}})} \geq \frac{p(x \mid \hat{\theta})}{p(x \mid \theta_{\text{target}})} = \exp\left[\tfrac{1}{2}\Delta\chi^2_0\right]$$



all toys will be sampled more often, high-$\Delta\chi^2$ exponentially more often

The solution: mixture of many θ

- What does high $\Delta\chi^2(\theta_{\text{true}})$ mean?
  → there exists a different value $\hat{\theta}$ from which one is more likely to obtain the data x than from θ_{true}

- Just use all toys sampled from many θ_{smpl} at once

  $$p_{\text{smpl}}(x) = p(x \mid \{\theta\}) := \frac{1}{S}\sum_{s=1}^{S} p(x \mid \theta_s)$$

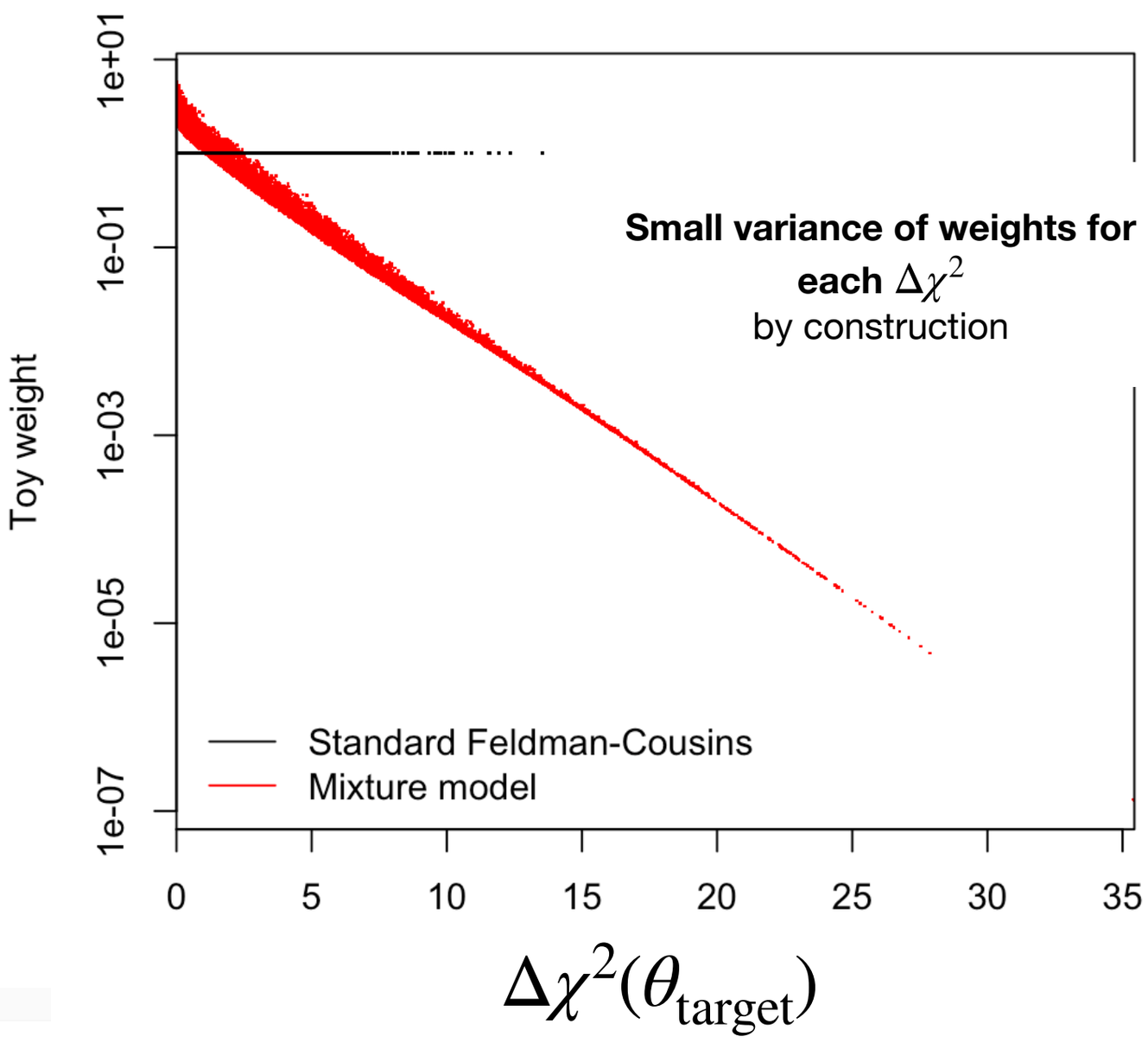
  (mixture model)

- As long as θ_s are close enough, there will always be some θ_s close enough to $\hat{\theta}$, such that the weights w will be bounded from above resulting in numerical stability (at least for the $\Delta\chi^2$ values we care about).

  $$w(x) \lesssim S \exp\left[-\frac{1}{2}\Delta\chi^2(\theta_{\text{target}})\right]$$

  with

  $$w(x) = \frac{p(x \mid \theta_{\text{target}})}{p(x \mid \{\theta\})}$$



The solution: mixture of many θ

- What does high $\Delta\chi^2(\theta_{\text{true}})$ mean?
  → there exists a different value $\hat{\theta}$ from which one is more likely to obtain the data x than from θ_{true}

- Just use all toys sampled from many θ_{smpl} at once

$$p_{\text{smpl}}(x) = p(x \mid \{\theta\}) := \frac{1}{S}\sum_{s=1}^{S} p(x \mid \theta_s)$$

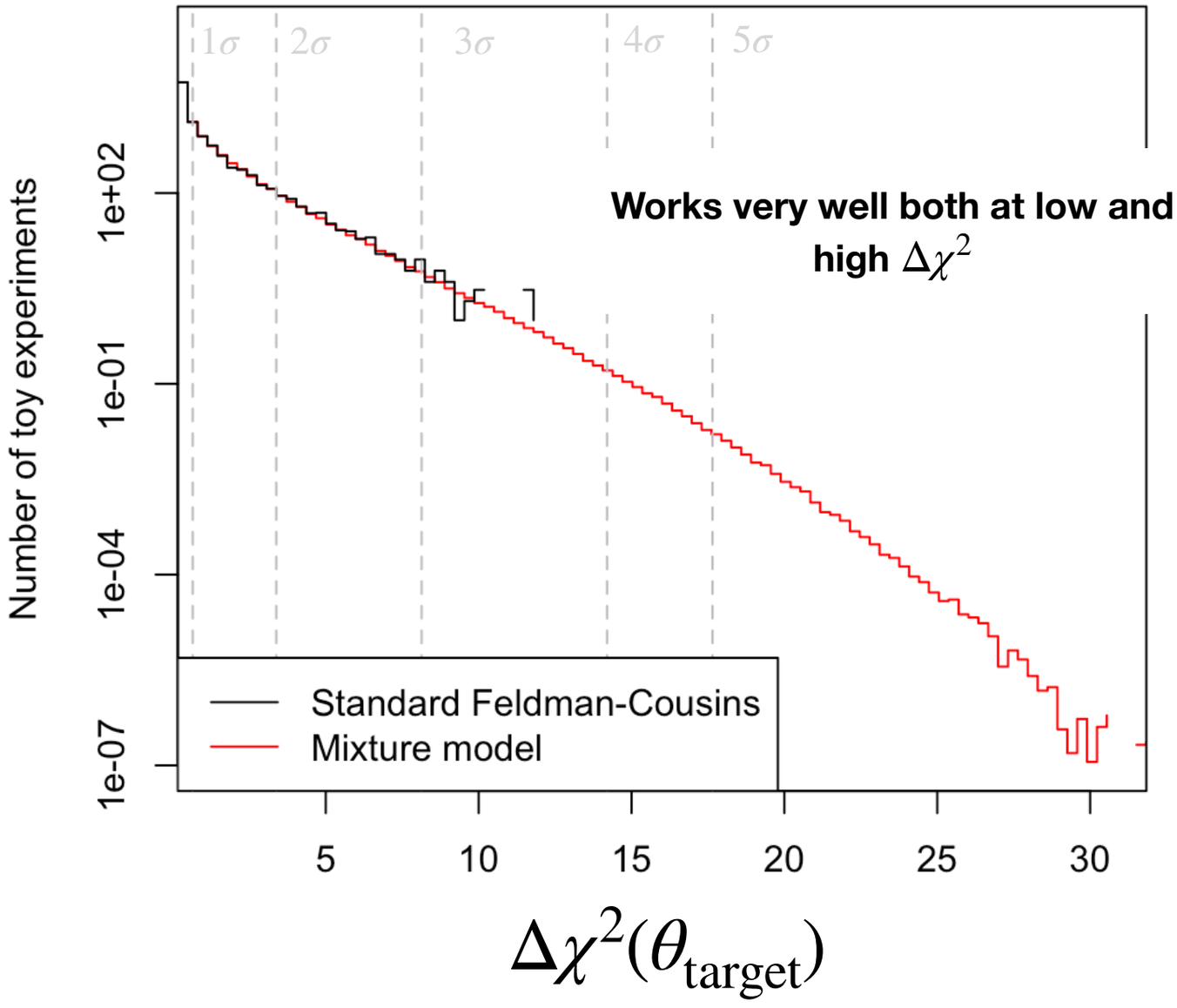
(mixture model)

- As long as θ_s are close enough, there will always be some θ_s close enough to $\hat{\theta}$, such that the weights w will be bounded from above resulting in numerical stability (at least for the $\Delta\chi^2$ values we care about).

$$w(x) \lesssim S \exp\left[-\tfrac{1}{2}\Delta\chi^2(\theta_{\text{target}})\right]$$

with

$$w(x) = \frac{p(x \mid \theta_{\text{target}})}{p(x \mid \{\theta\})}$$

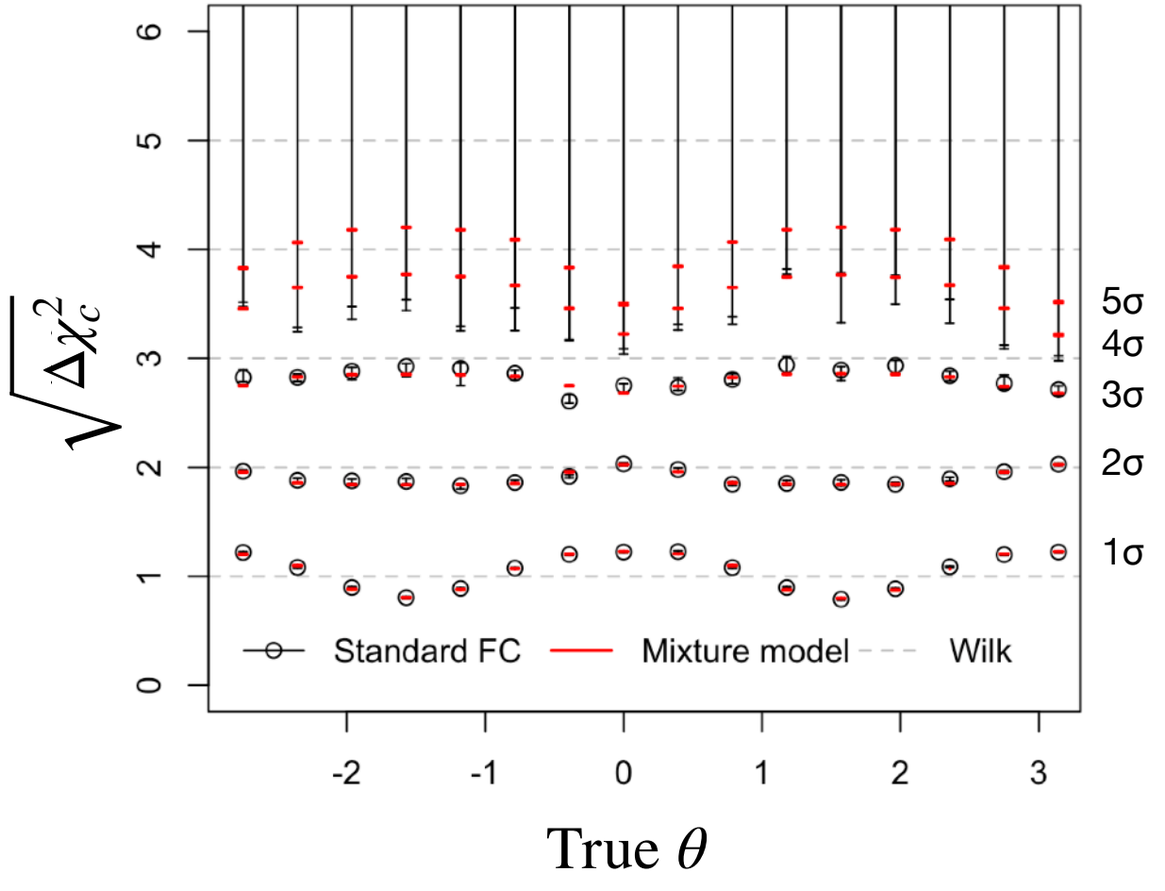


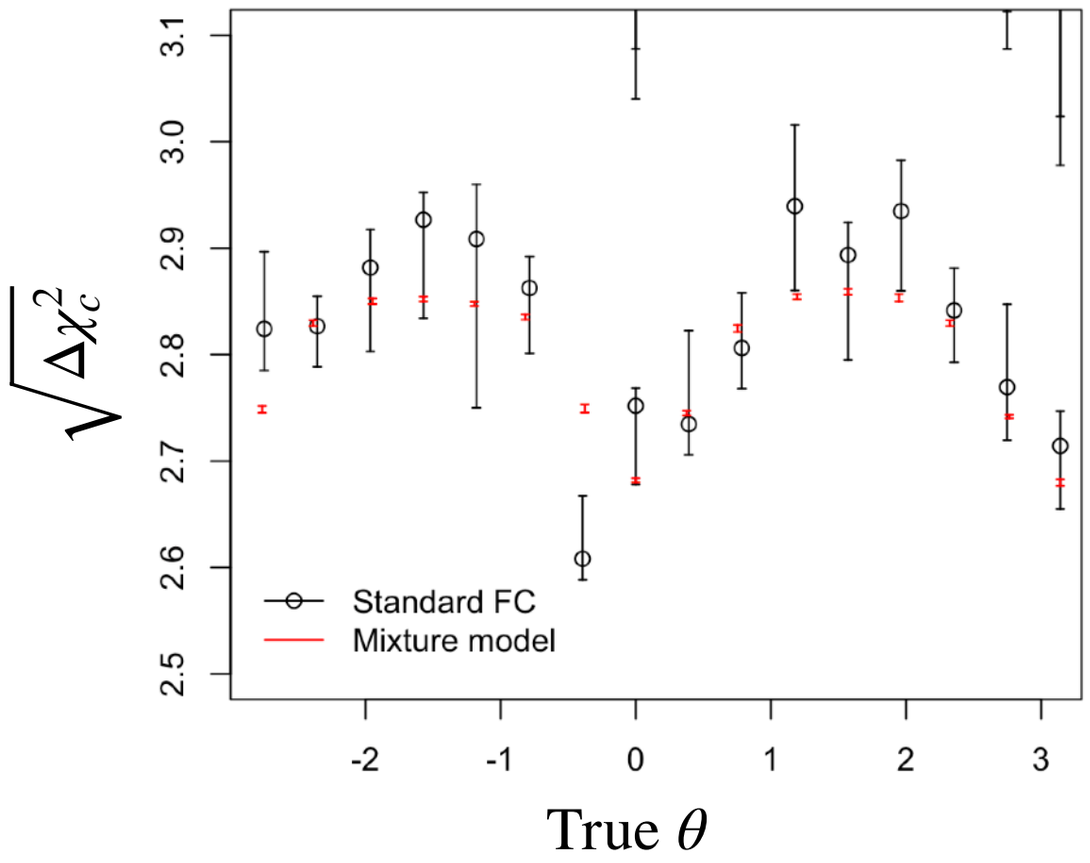
Critical value distributions

- The same toys are used to produce both results!

- Errors for critical values significantly reduced (e.g. 3σ)

- 5σ were impossible to estimate before, now easy





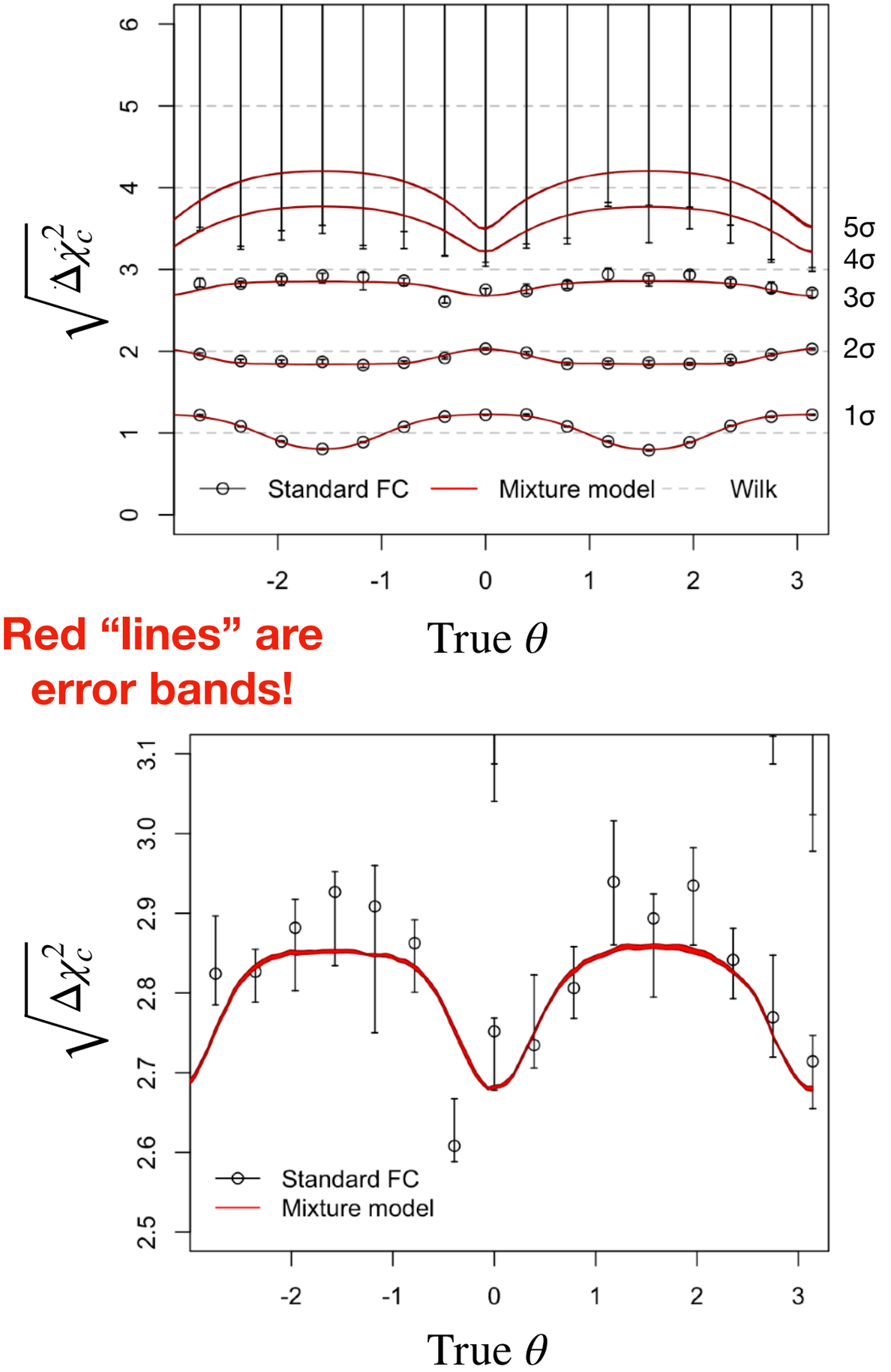
Critical value distributions

- The same toys are used to produce both results!

- Errors for critical values significantly reduced (e.g. 3σ)

- 5σ were impossible to estimate before, now easy

- Bonus: can now interpolate critical values for free!



Red "lines" are error bands!

Nuisance parameters

- Feldman-Cousins paper itself does not deal with presence of nuisance parameters, various approaches exist

- As prior distribution (**prior Highland-Cousins**)

  Just define $p(x \mid \theta) := \int \mathrm{d}\eta \, p(\eta) \, p(x \mid \theta, \eta)$

  and use this both for fitting and throwing toys, i.e. fit by **marginalizing**.
  In most cases using profile likelihood is similar to marginal likelihood so should be ok for using in fit too (with constraint term corresponding to prior).

  Performance improvement same.

  If systematics are not well constrained by data,
  can also compute toy weight at fixed η instead of marginalizing, which is much cheaper.
  i.e. $w(x, \eta_{\text{true}}) = \dfrac{p(x \mid \theta_{\text{target}}, \eta_{\text{true}})}{p(x \mid \{\theta\}, \eta_{\text{true}})}$ instead of $w(x) = \dfrac{p(x \mid \theta_{\text{target}})}{p(x \mid \{\theta\})}$

- For prior unconstrained parameters and parameters that get very constrained by the data this is not ideal. Use **posterior Highland-Cousins** conditioned on the true value of the parameter of interest θ:

  i.e. define toy distribution by $p(x \mid \theta) := \int \mathrm{d}\eta \, p(\eta \mid x_{\text{data}}, \theta) \, p(x \mid \theta, \eta)$

  and fit by **profiling**. Here the constraint term should be thrown for each toy experiment (included in x) as though it is external data.

  In this case we get the same performance improvement *if* for fixed θ the η-posterior is **sufficiently gaussian**.

- I do not know yet, if one can prove any properties if η-posterior is not sufficiently gaussian, and method may not be directly applicable. However, in this case the Feldman-Cousins method will be dependent on the nuisance parameter treatment choice (similar to priors in Bayesian analysis) so more careful analysis would be necessary anyway. If the number of such non-gaussian nuisances is small, they could be included as part of the parameters of interest.

Summary

- Neutrino oscillation experiments aiming to discover CP violation in leptonic sector

- Frequentist analysis of neutrino oscillation requires expensive Feldman-Cousins procedure because Wilk's theorem cannot be used

- Proposed a simple technique that efficiently generates high-$\Delta\chi^2$ toys, resulting in exponential reduction of uncertainty for high-$\Delta\chi^2$

  + provides correct interpolation of critical values for free

- Hopefully this can help many experiments, not only ν -osc.

- Exploring various other techniques as well, hopefully Feldman-Cousins will not be considered an expensive calculation soon

"Simply use a mixture of toys generated at many θ values and reweight"

Similar to "multiple histogram reweighting" in statistical mechanics / lattice QCD calculations see e.g. Kari Rummukainen, "Monte Carlo simulation methods" lecture notes on reweighting

I think some Higgs p-value studies (by K. Cranmer's group) also used similar methods.

to be posted on arXiv soon!