Neural Simulation-Based Inference

Aishik Ghosh, Eddie McGrady, Tae Hyun Park

ML4FP School 2026
2 June 2026

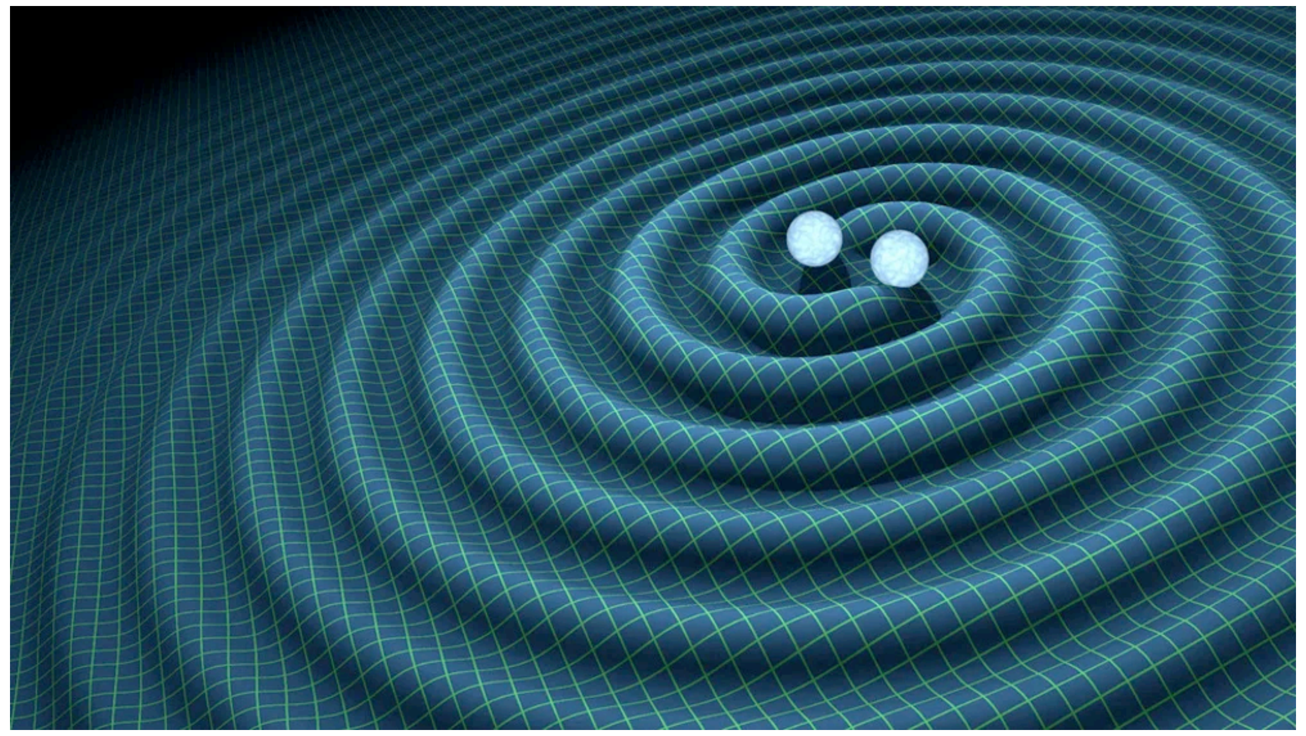
Georgia Institute of Technology

Some new statistics concepts for some of you, ask questions!

Prof-Aishik-Ghosh
aishikghosh.bsky.social
@Aishik_Ghosh_

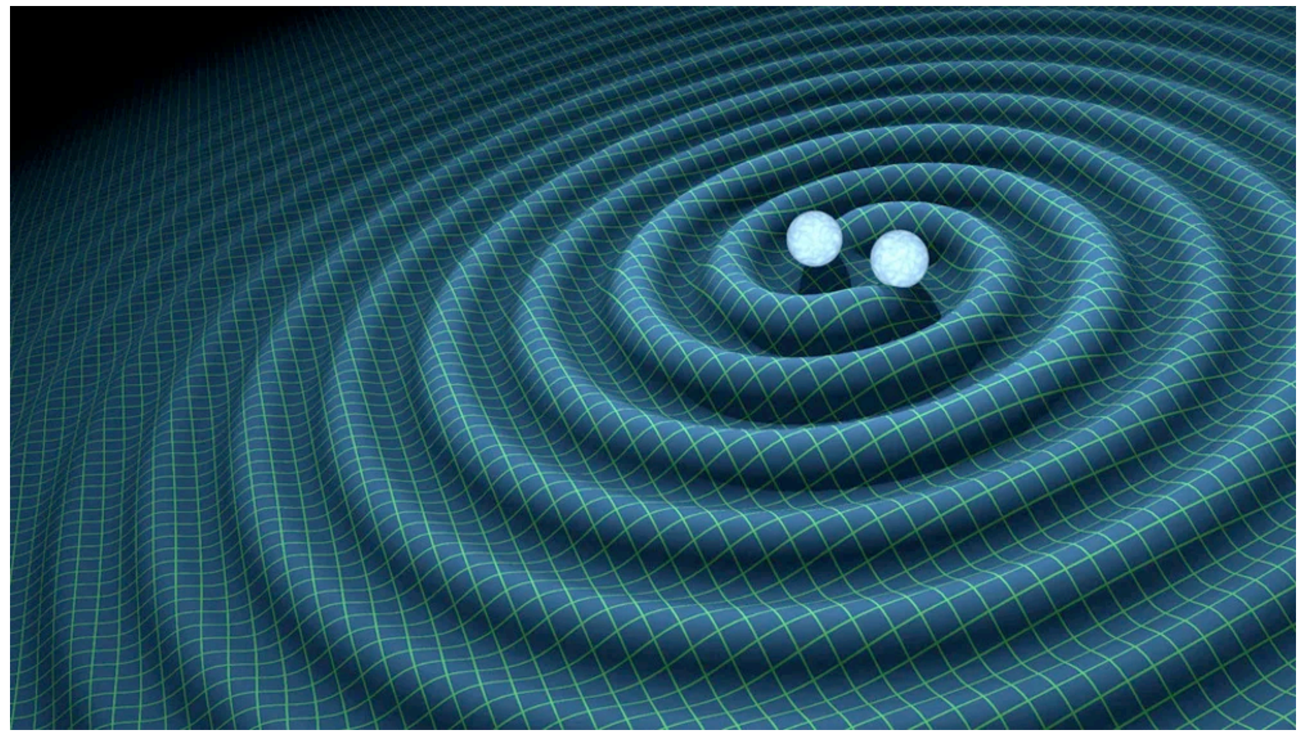
Gravitational Wave discovery (2016)

Astrophysics:

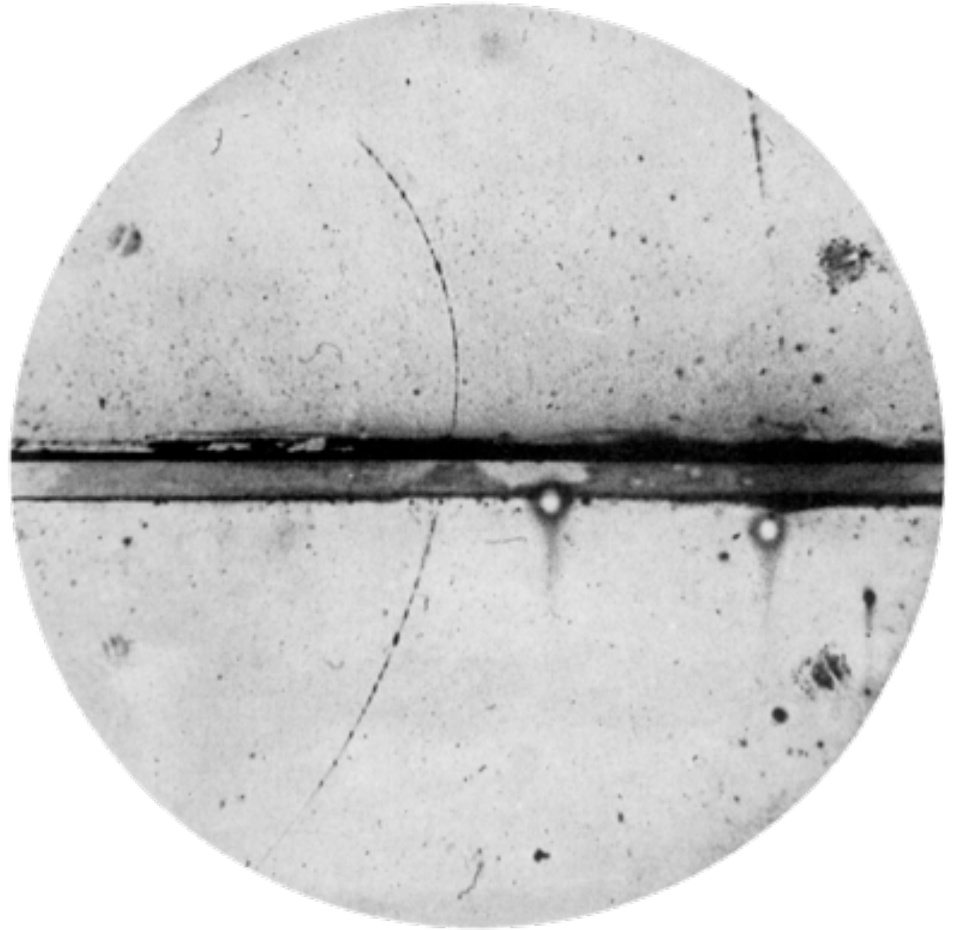


Gravitational Wave discovery (2016)

Astrophysics:



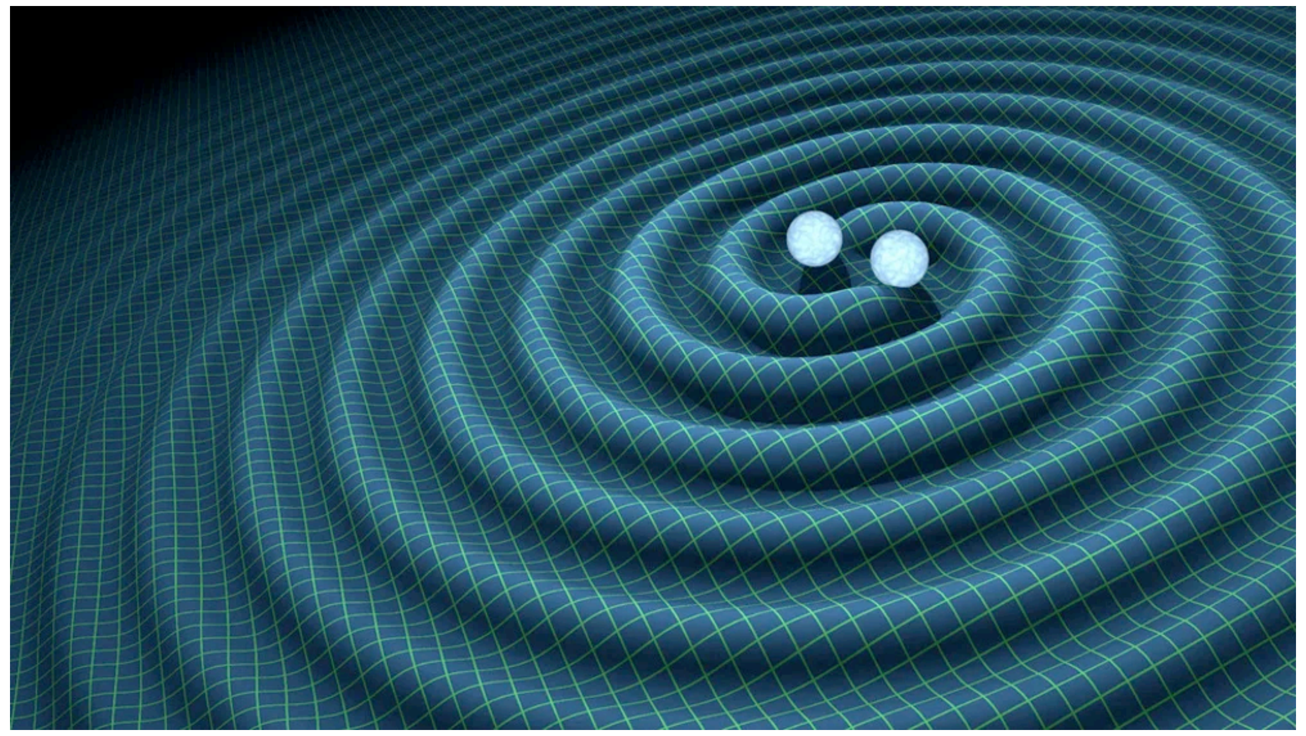
Particle physics:



Positron discovery (1930s)

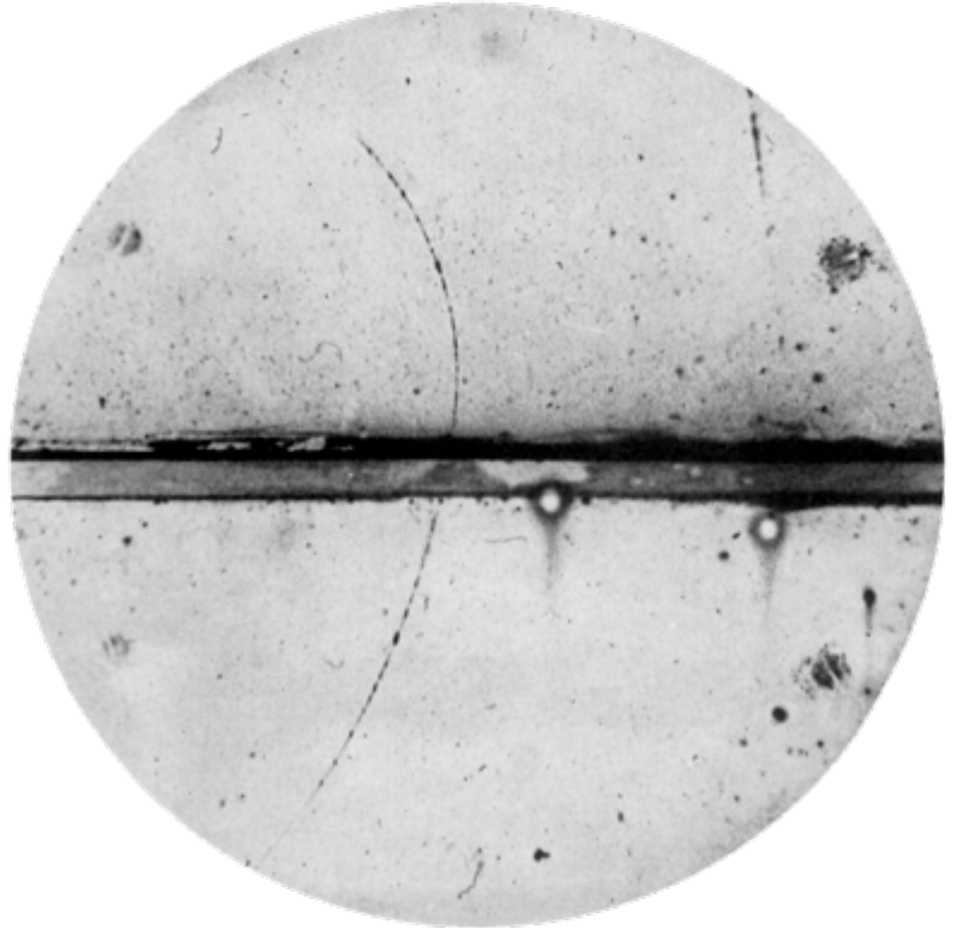
Gravitational Wave discovery (2016)

Astrophysics:



Single event

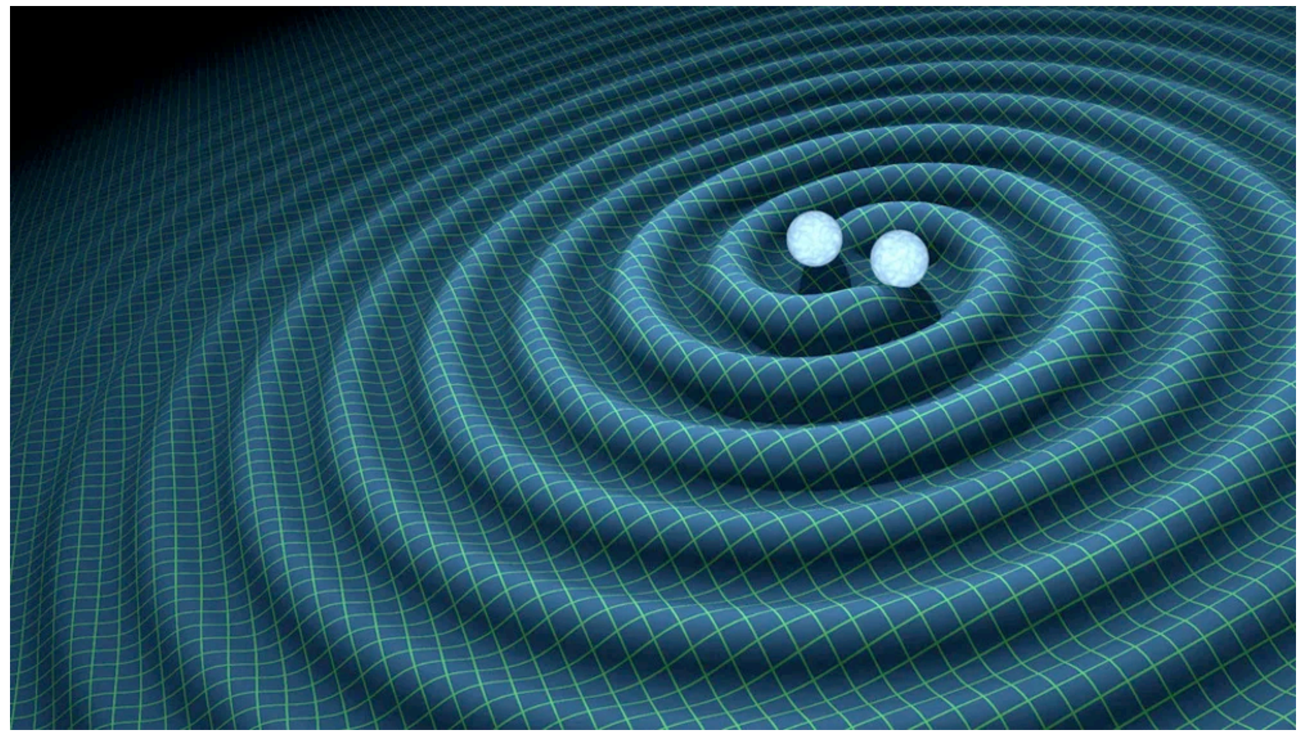
Particle physics:



Positron discovery (1930s)

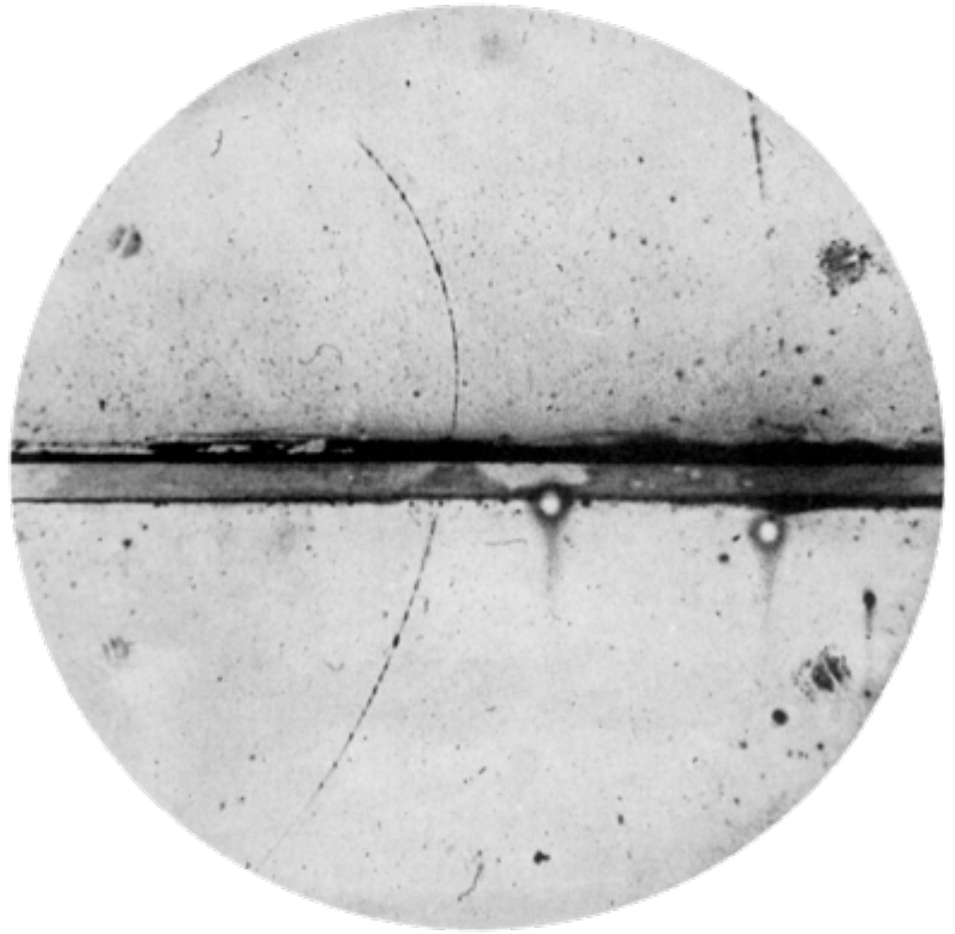
Gravitational Wave discovery (2016)

Astrophysics:

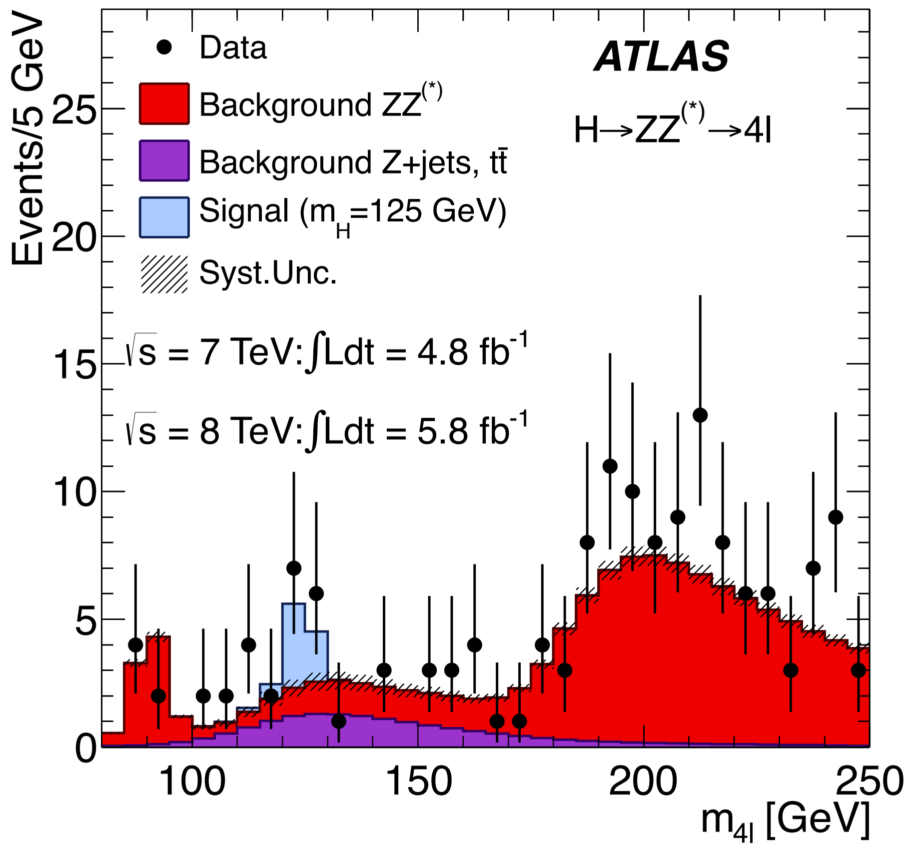


Single event — Collection of events

Particle physics:



Positron discovery (1930s)

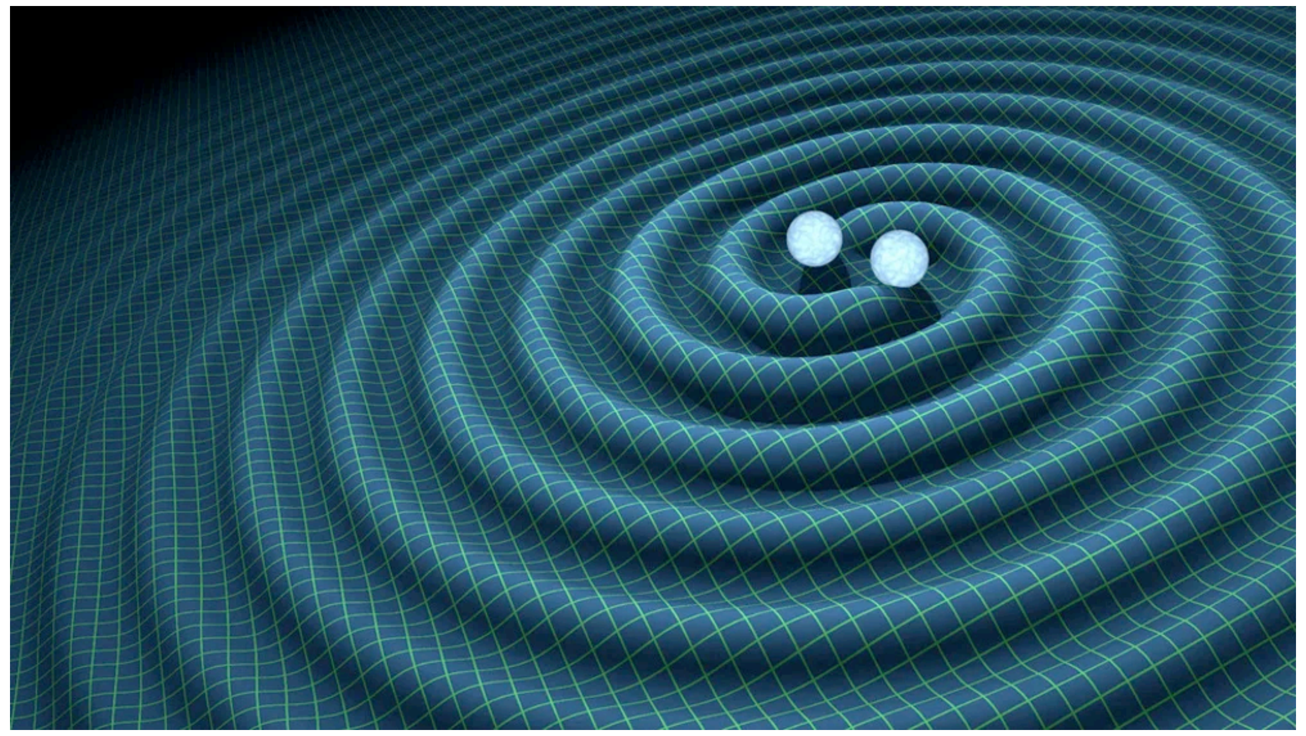


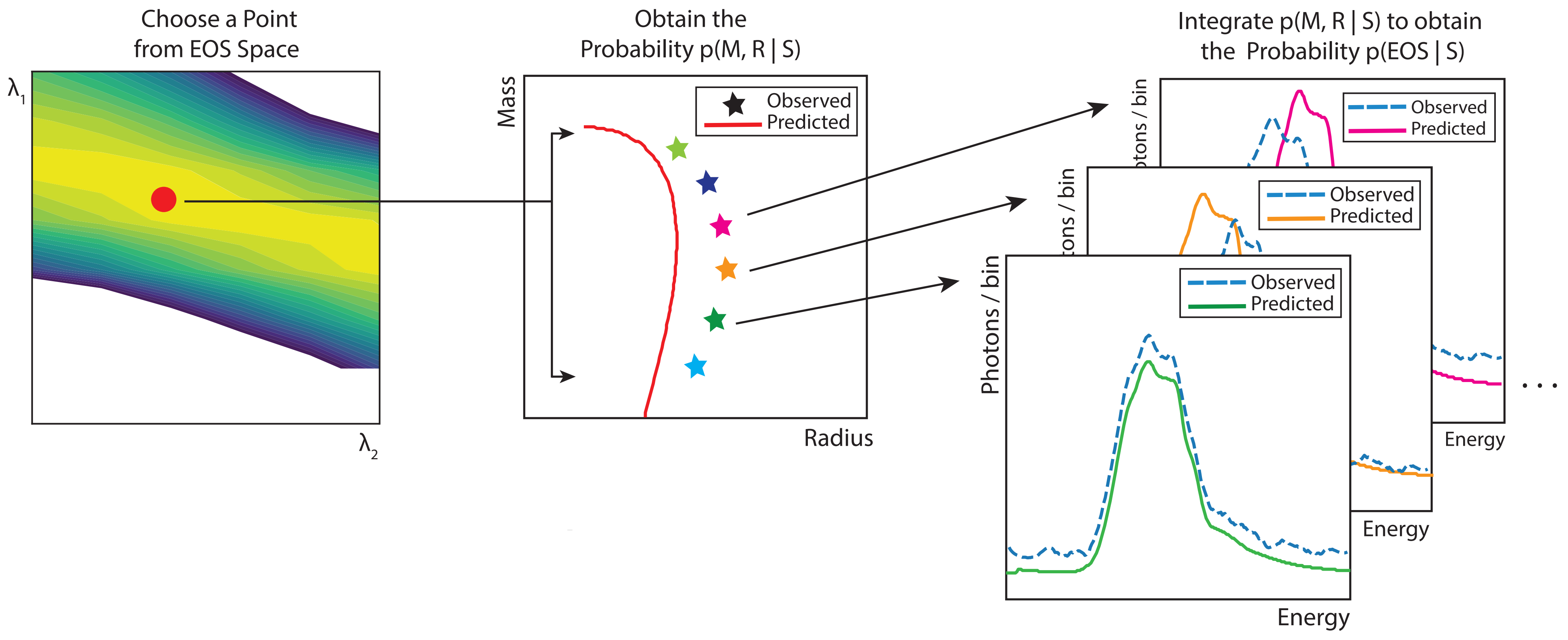
Higgs boson discovery (2012)

Gravitational Wave discovery (2016)

Probing the interior of neutron stars

Astrophysics:

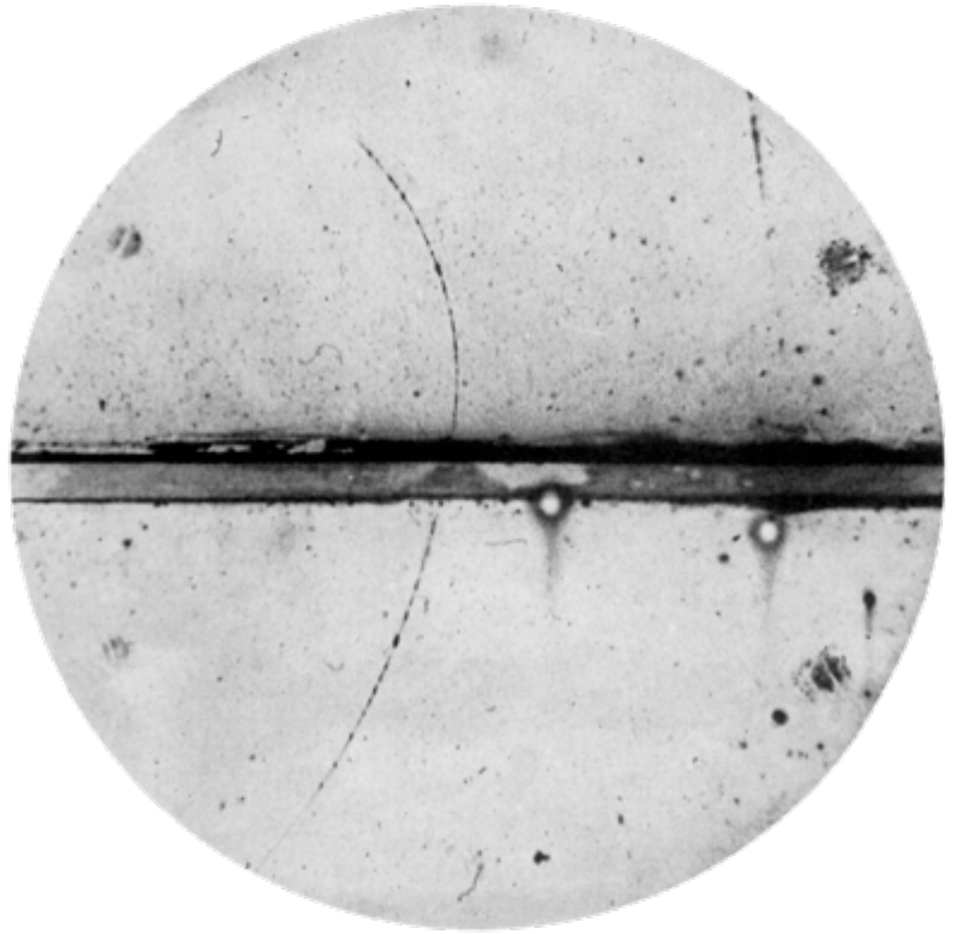


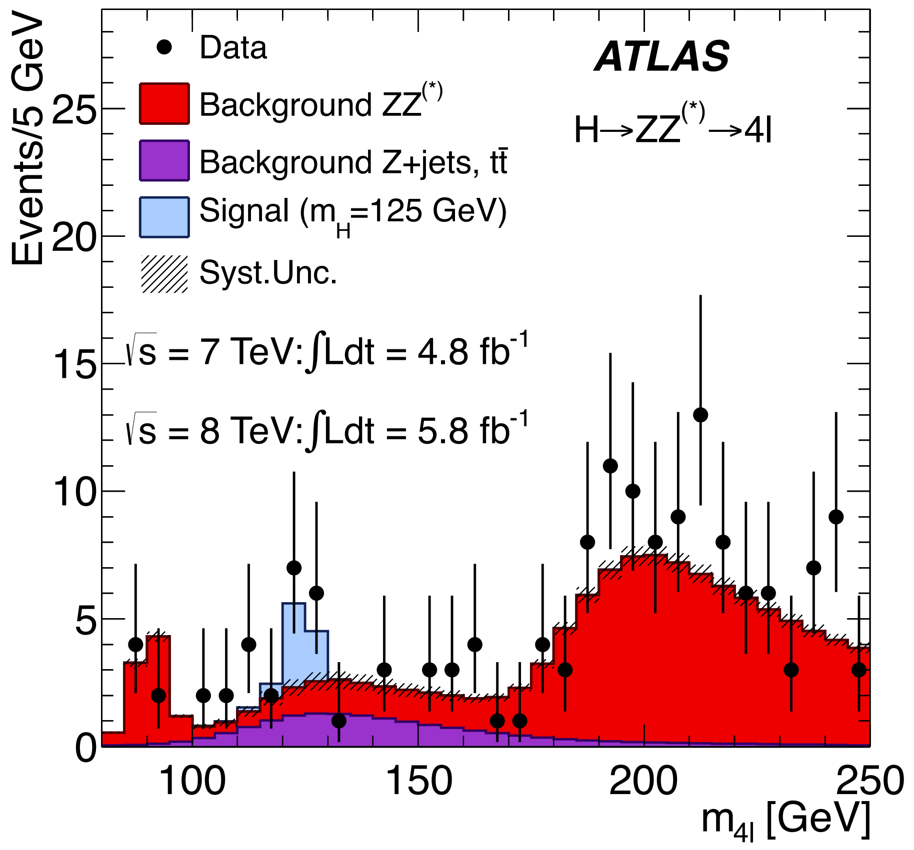


JCAP12(2023)022: Farrell et al. (incl. Ghosh)

Single event — Collection of events

Particle physics:





Positron discovery (1930s)

Higgs boson discovery (2012)

Statistical inference of physics parameters

mH = 125.25 ± 0.17 GeV

Statistical inference of physics parameters

mH = 125.25 ± 0.17 GeV

Statistical inference of physics parameters

$$mH = 125.25 \pm 0.17 \text{ GeV}$$

"How sure am I ? How can I reduce my uncertainty ?"

Statistical inference of physics parameters

mH = 125.25 ± 0.17 GeV

Statistical uncertainties, but also detector systematic, theory systematic, MC stat, ...

“How sure am I ? How can I reduce my uncertainty ?”

An experimentalist thinks about ‘confidence intervals’

Generate data at a fancy experiment

Video: CERN

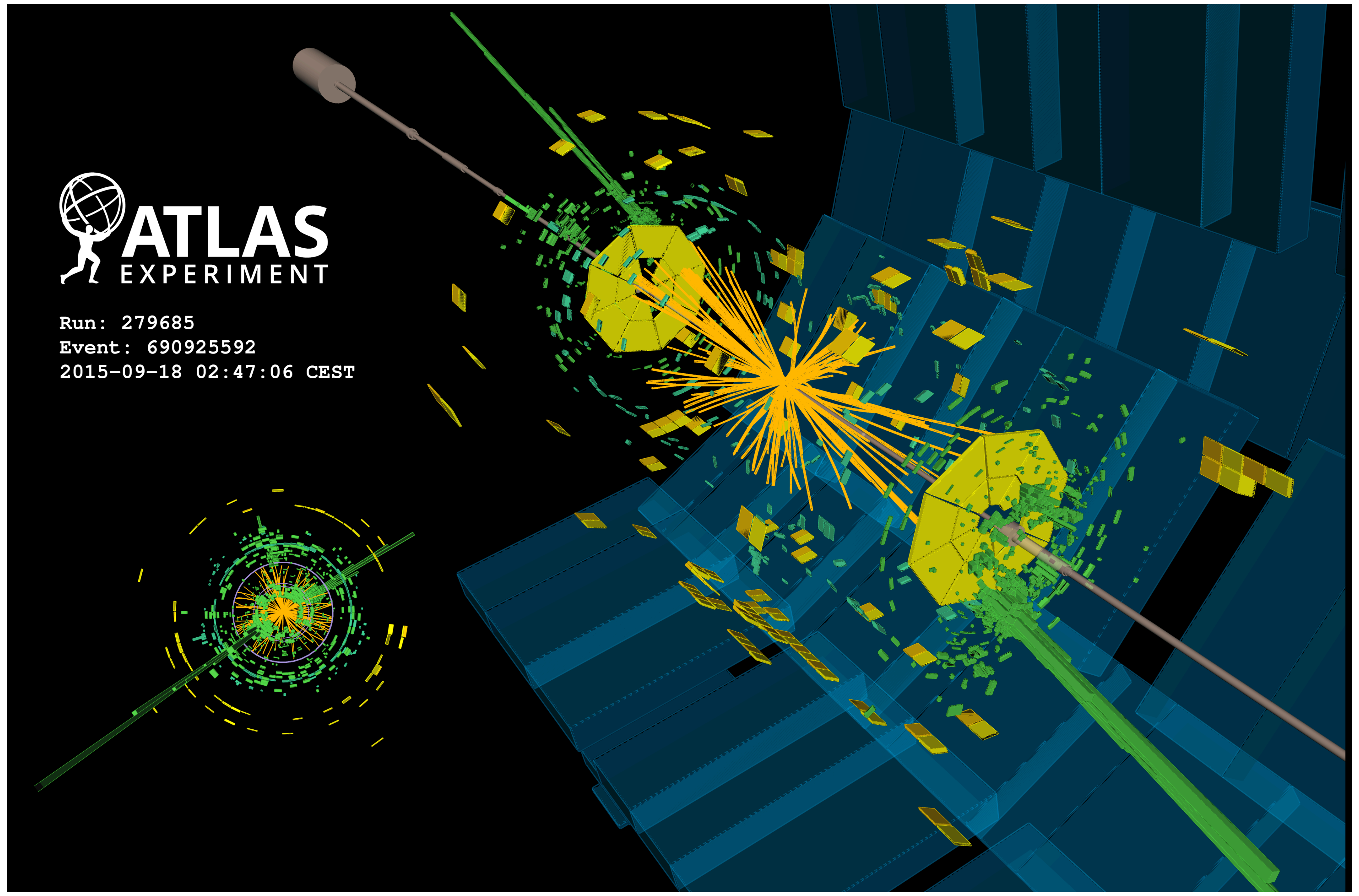
Generate data at a fancy experiment

Video: CERN

Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

- Reconstruction pipeline, event selection
- Design one summary variable
    - Compression: O(100 million) → 1
- Build a 1-D histogram
- Estimate probability densities with histograms

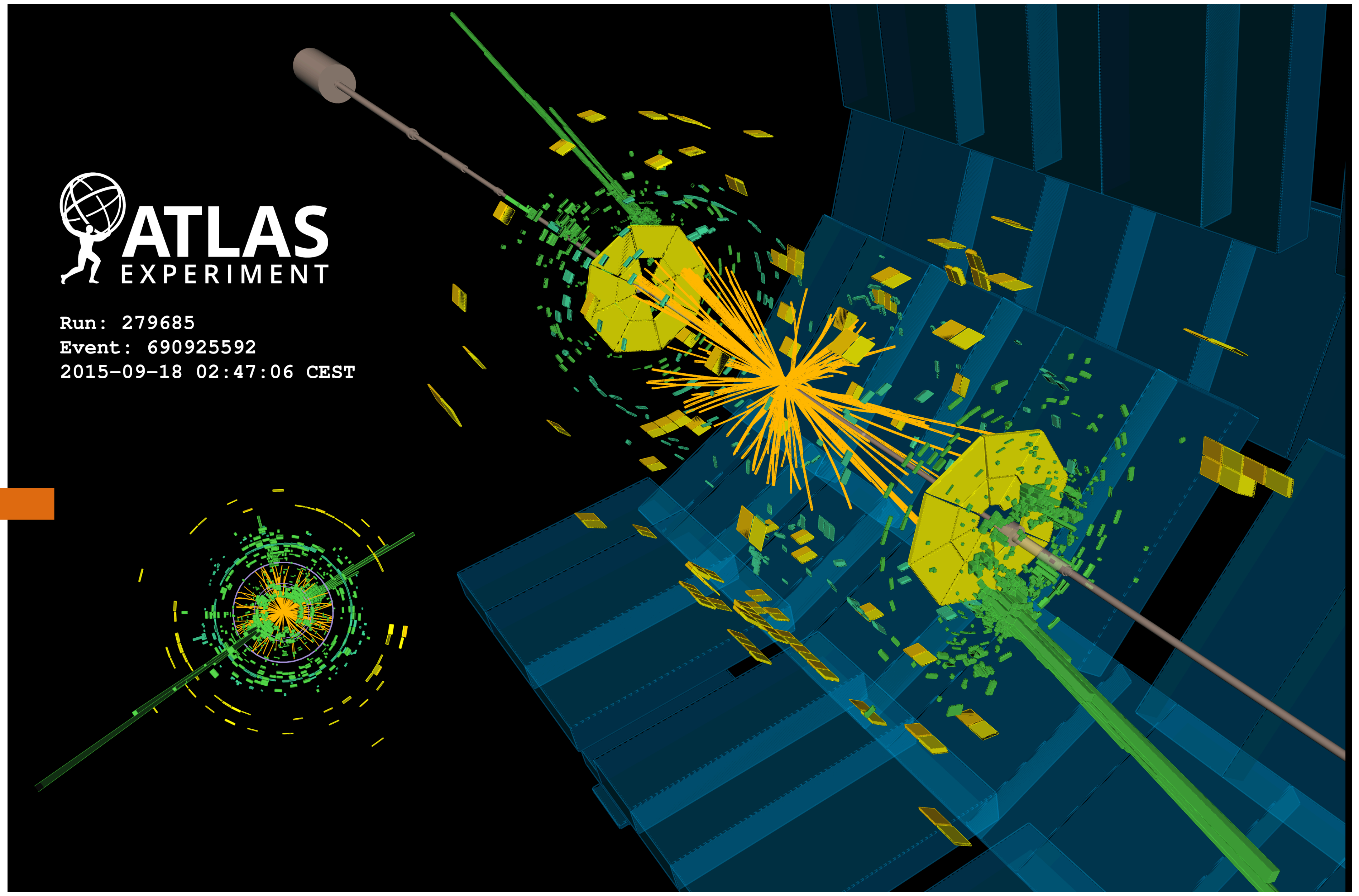


Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

- Reconstruction pipeline, event selection
- Design one summary variable
    - Compression: O(100 million) → 1
- Build a 1-D histogram
- Estimate probability densities with histograms

1 number

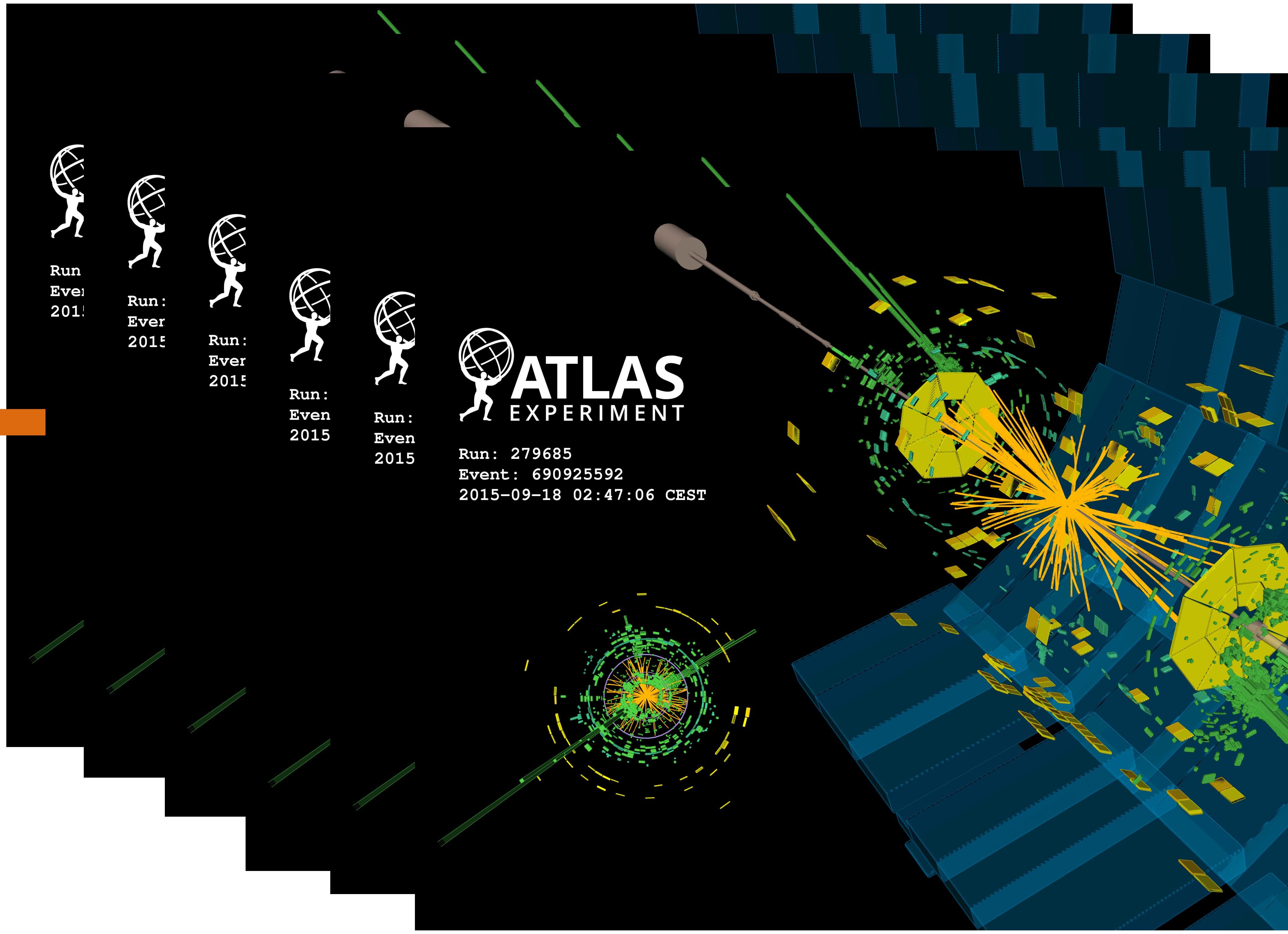


Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

- Reconstruction pipeline, event selection
- Design one summary variable
  - Compression: O(100 million) → 1
- Build a 1-D histogram
- Estimate probability densities with histograms

Thousands of collision events



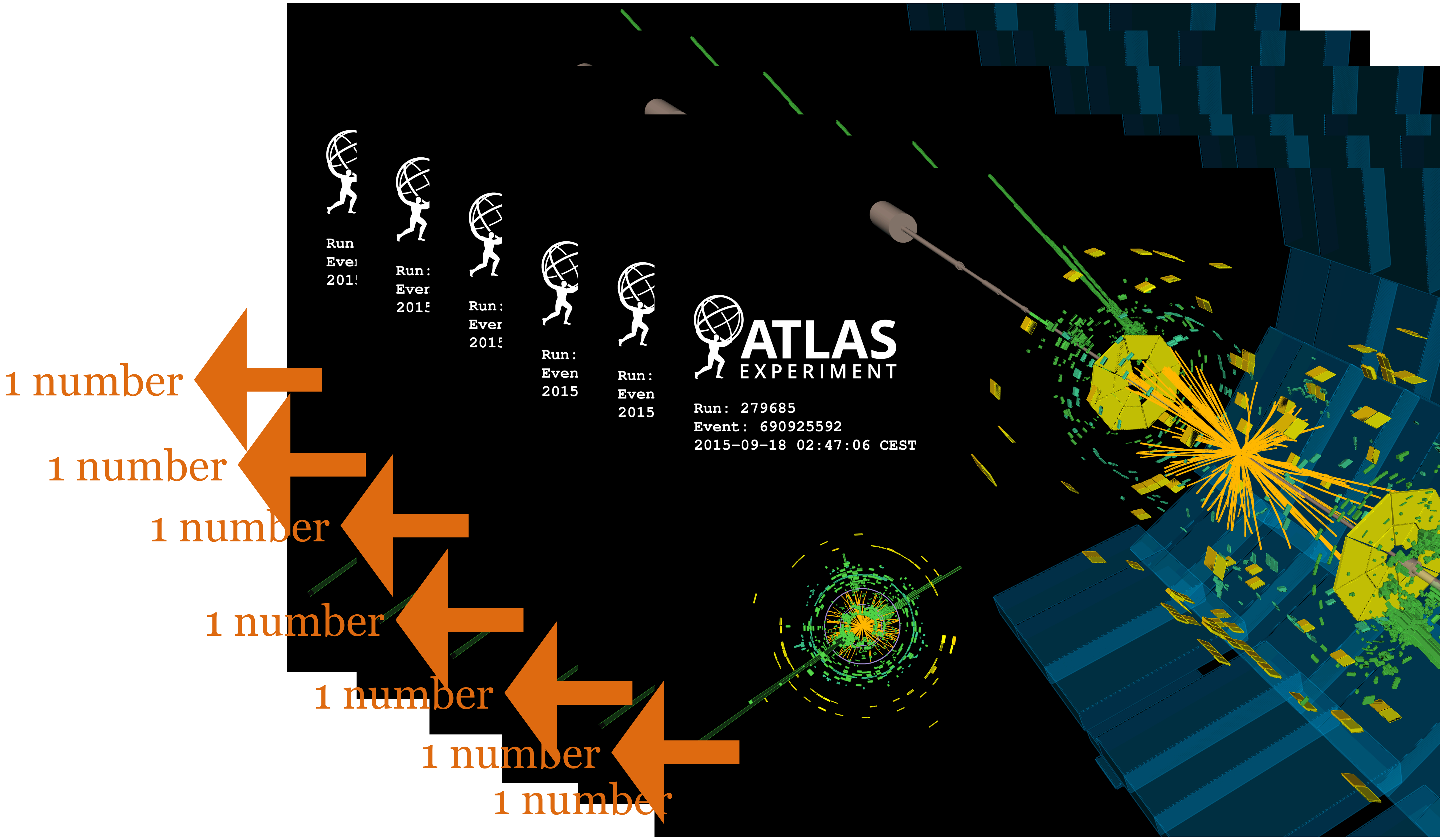
1 number

Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

- Reconstruction pipeline, event selection
- Design one summary variable
    - Compression: O(100 million) → 1
- Build a 1-D histogram
- Estimate probability densities with histograms

Thousands of collision events

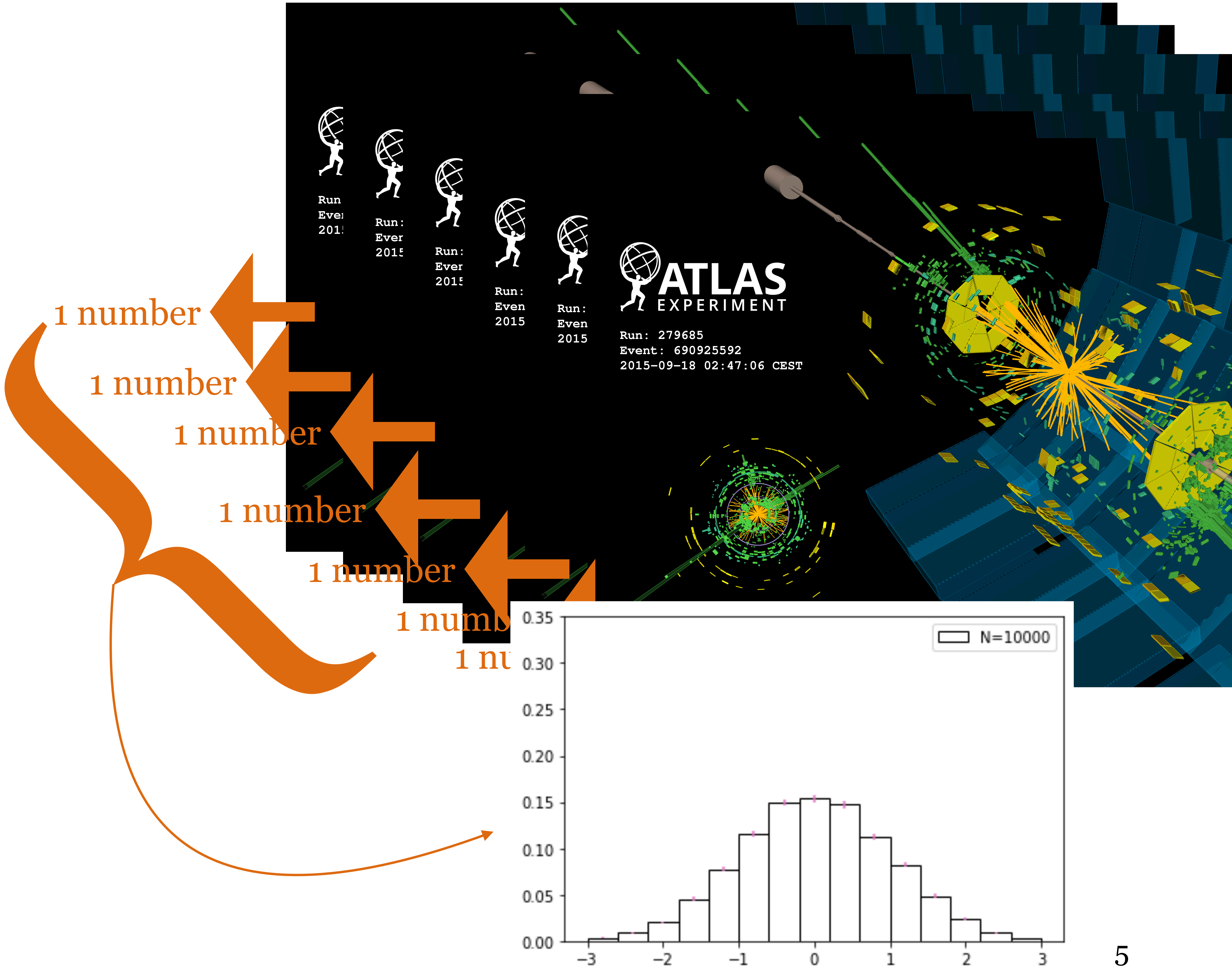


Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

- Reconstruction pipeline, event selection
- Design one summary variable
  - Compression: O(100 million) → 1
- Build a 1-D histogram
- Estimate probability densities with histograms

Thousands of collision events



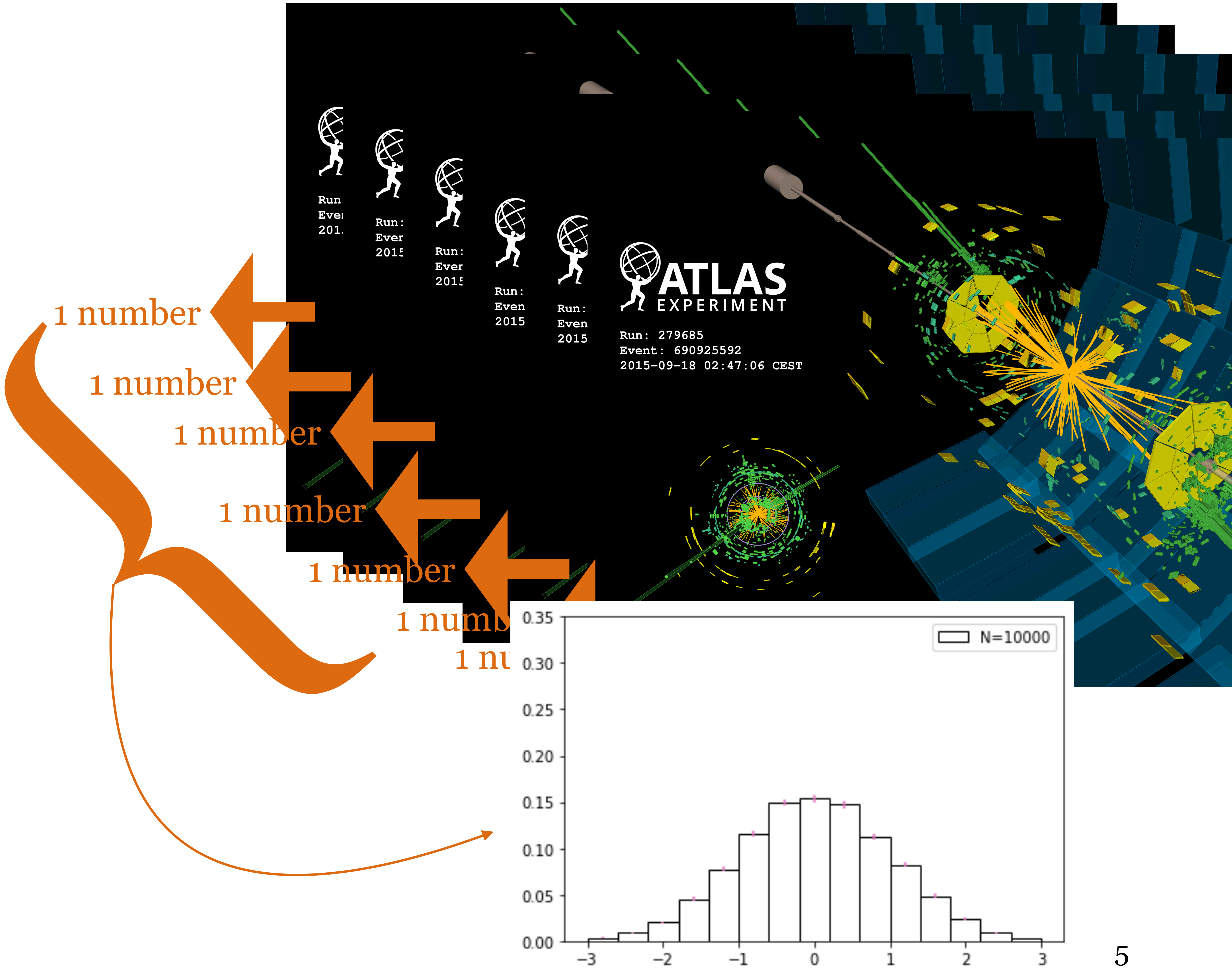
Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

- Reconstruction pipeline, event selection
- Design one summary variable
    - Compression: O(100 million) → 1
- Build a 1-D histogram
- Estimate probability densities with histograms

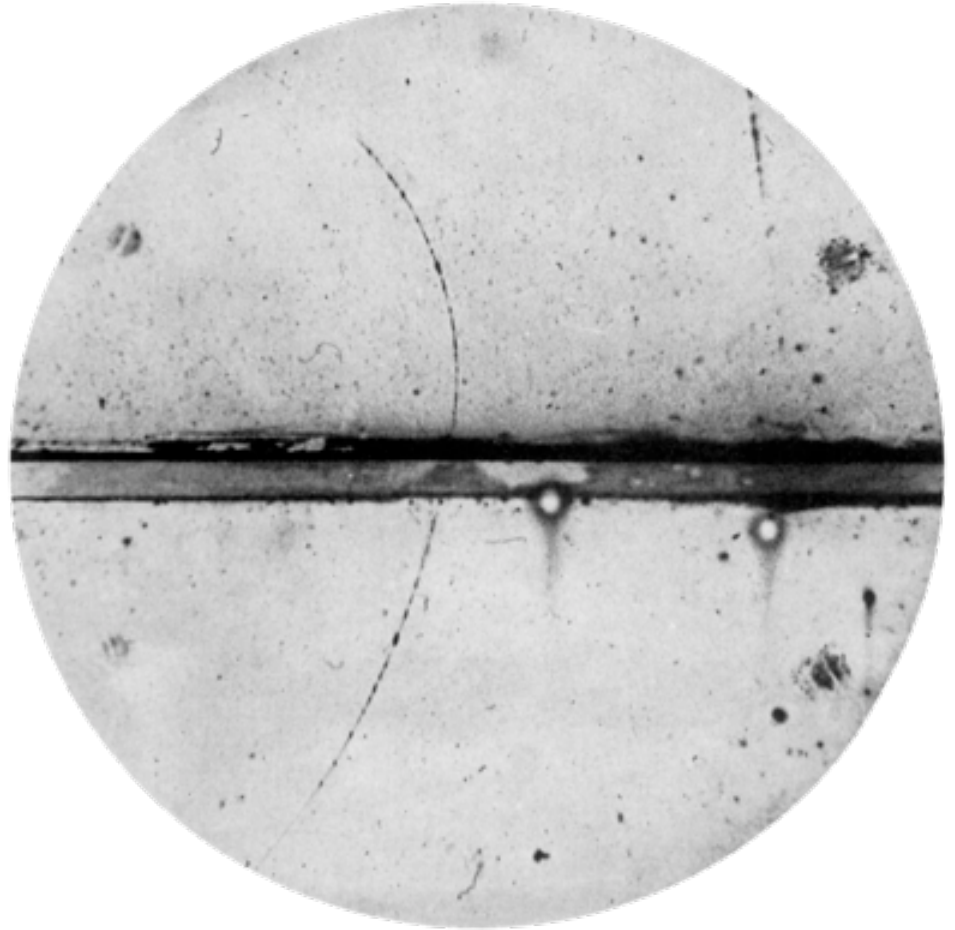
Why histogram help calculate probabilities?

Thousands of collision events



Positron discovery (1930s)

Top quark discovery (1990s)



Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}

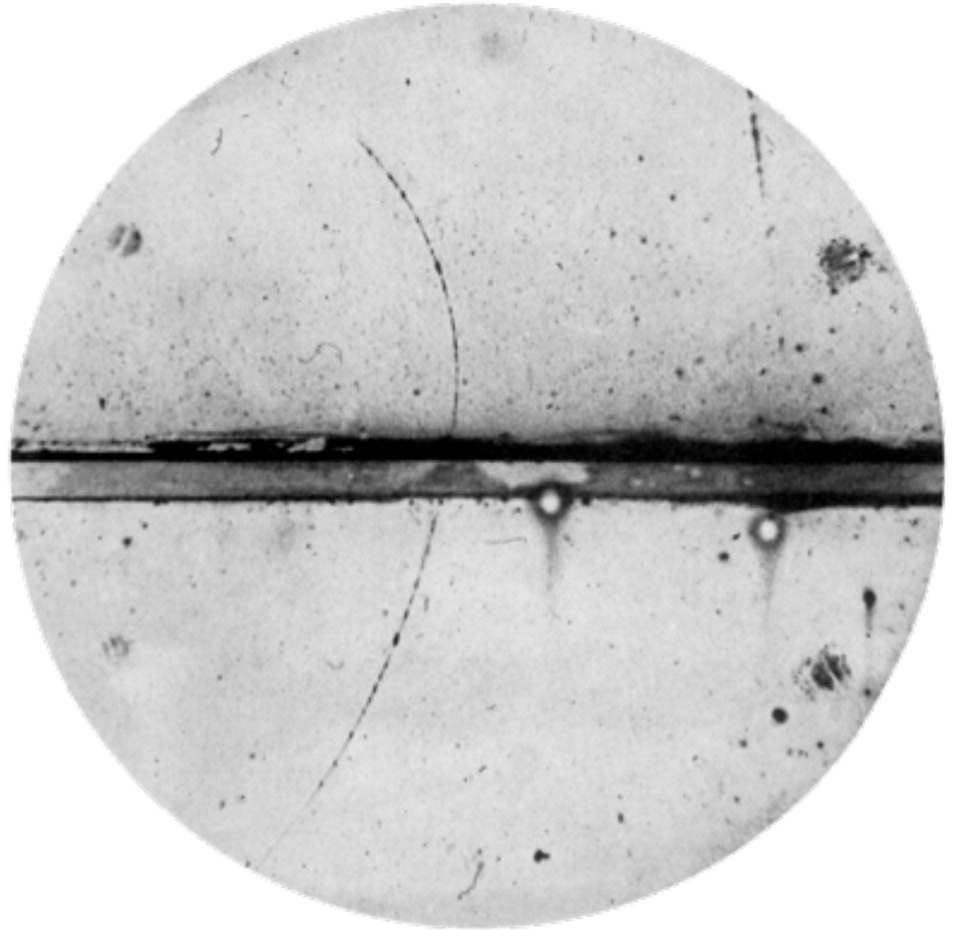
Single event

Multiple events:
Cut-and-count

Image: Wikipedia / PhysRev.43.491

CDF Collaboration: arXiv:9503002

Positron discovery (1930s)



Single event

Image: Wikipedia / PhysRev.43.491

Top quark discovery (1990s)

Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}

Multiple events:
Cut-and-count

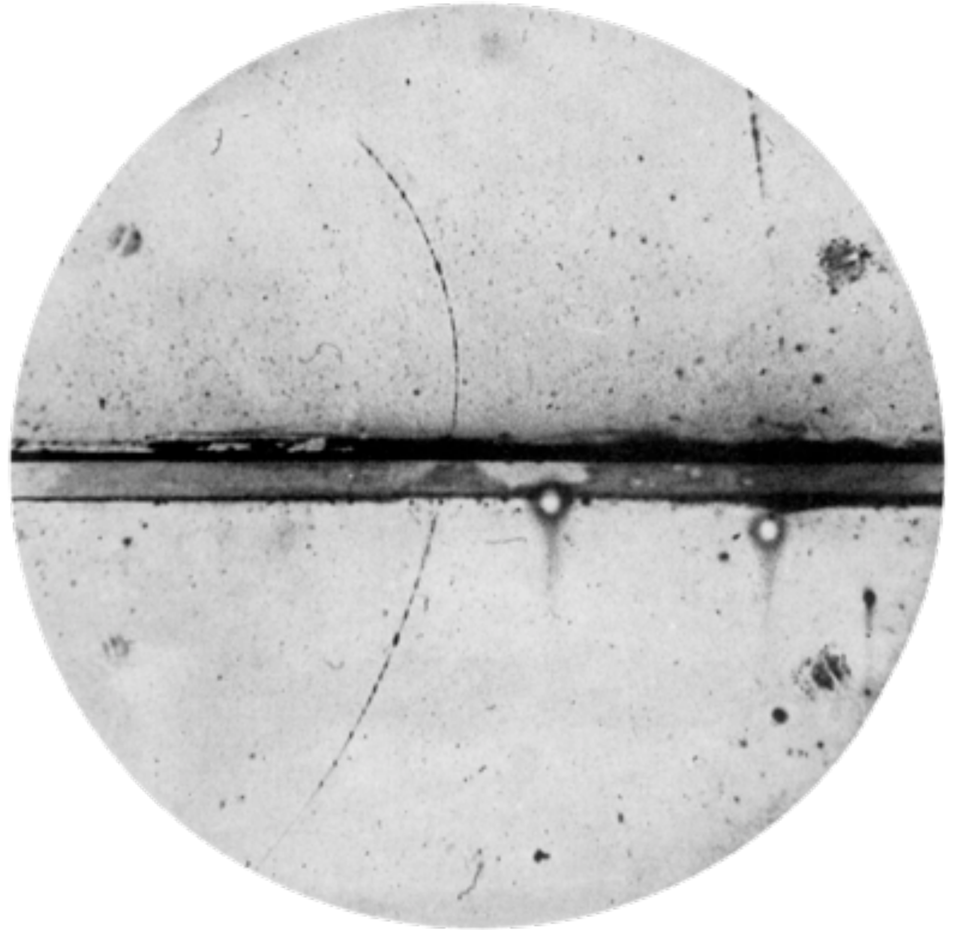
CDF Collaboration: arXiv:9503002

What is the probability of seeing 27 events if my null hypothesis predicts 7?

Positron discovery (1930s)

Top quark discovery (1990s)

What is the probability of seeing 27 events if my null hypothesis predicts 7?



Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}

$$p(n_{obs}\,|\,n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

Counting experiment → Poisson distribution

Single event

Multiple events:
Cut-and-count

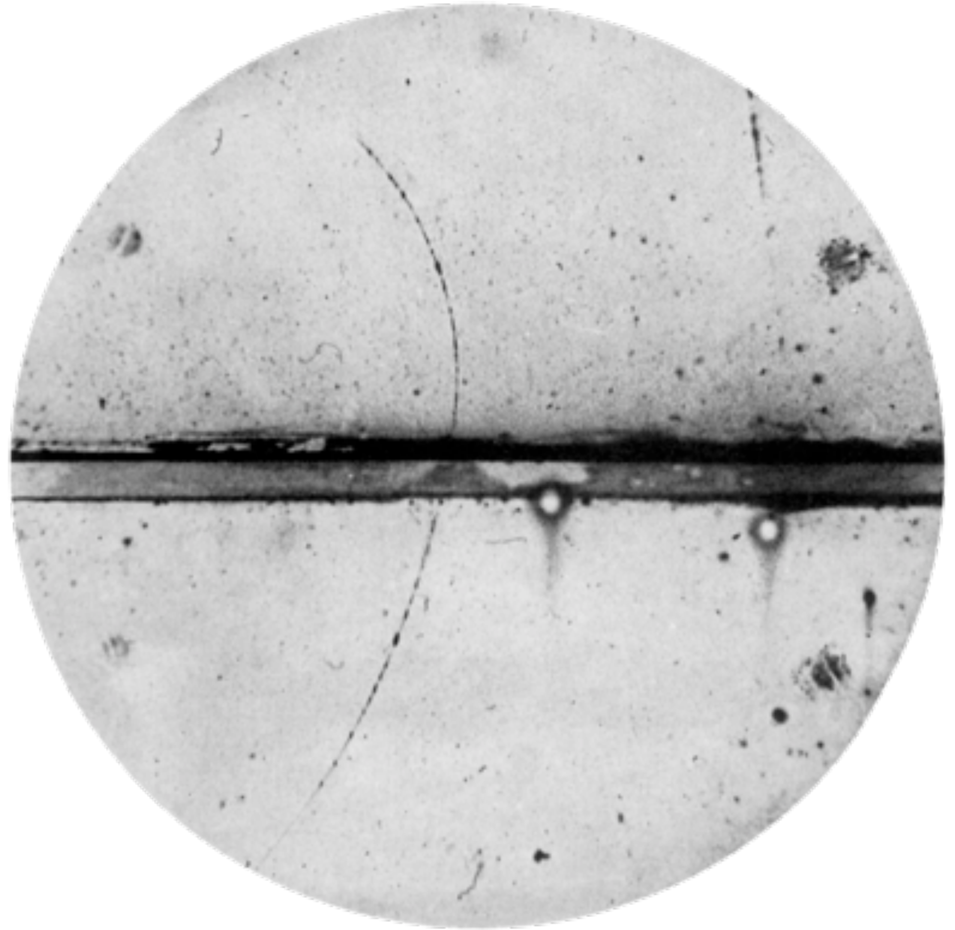
Image: Wikipedia / PhysRev.43.491

CDF Collaboration: arXiv:9503002

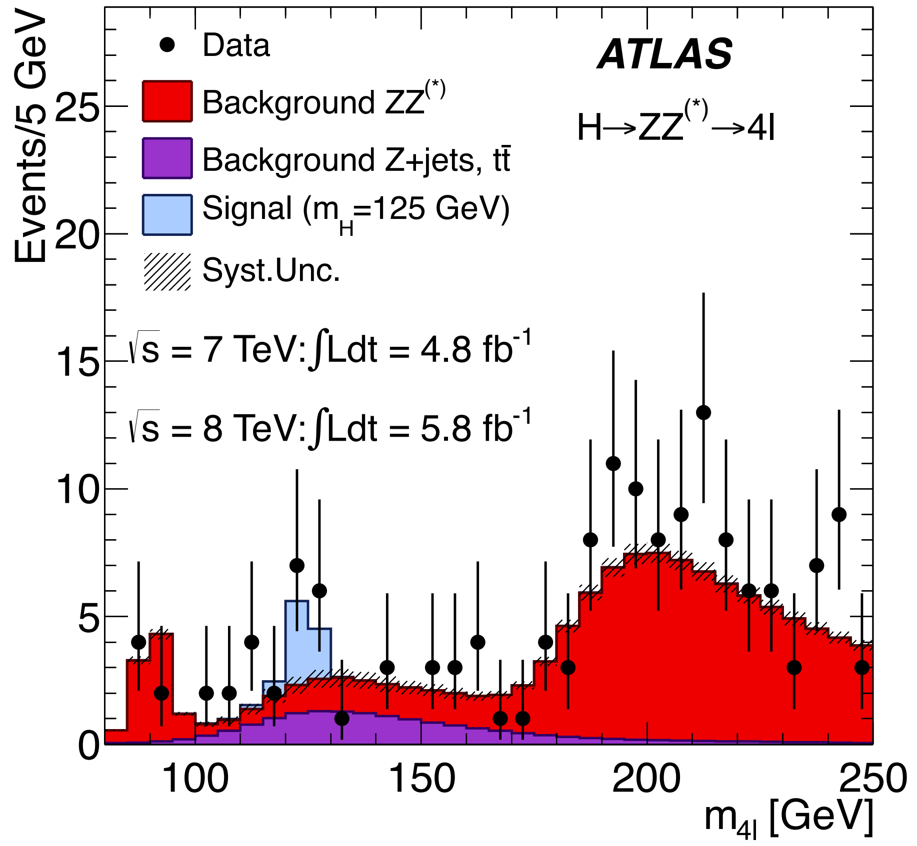
Positron discovery (1930s)

Top quark discovery (1990s)

Higgs boson discovery (2010s)



Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}



$$p_{bin}(n_{obs} \,|\, n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

Histogram bins just a bunch of independent counting experiments

Single event

Multiple events: Cut-and-count

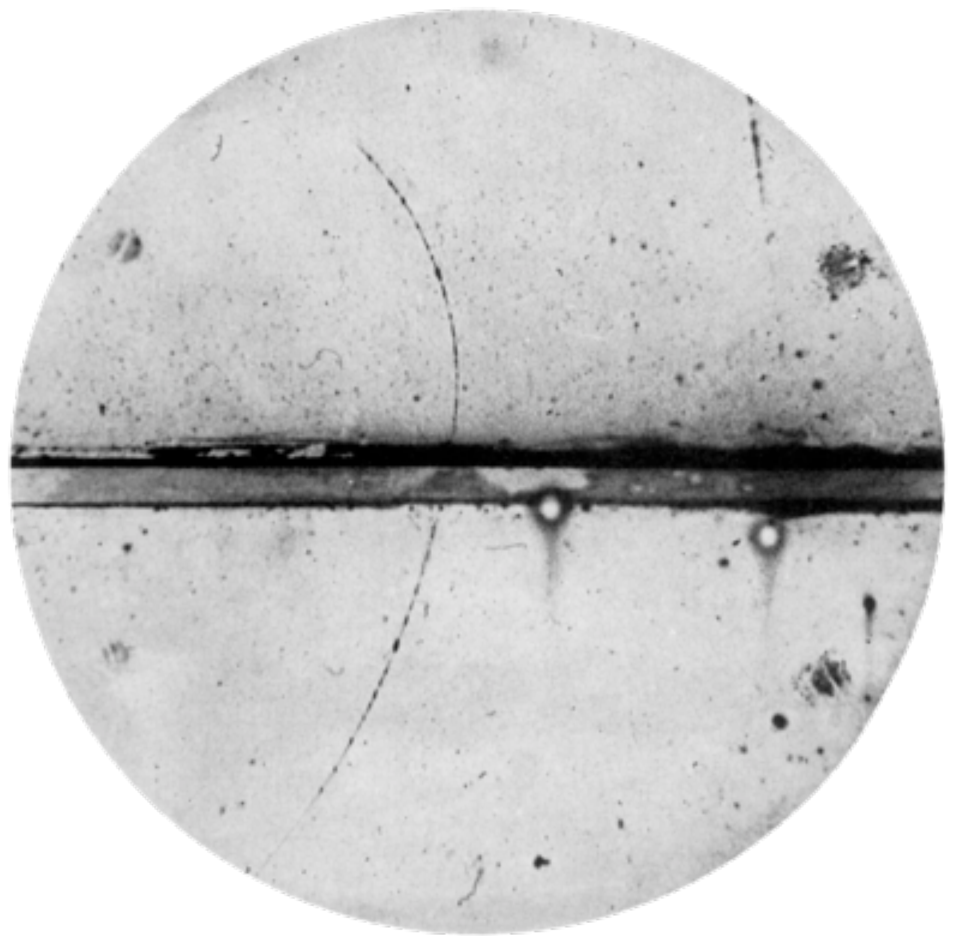
Shape information: Histogram

Image: Wikipedia / PhysRev.43.491

CDF Collaboration: arXiv:9503002

ATLAS Collaboration: arXiv:1207.7214

Positron discovery (1930s)



Single event

Image: Wikipedia / PhysRev.43.491

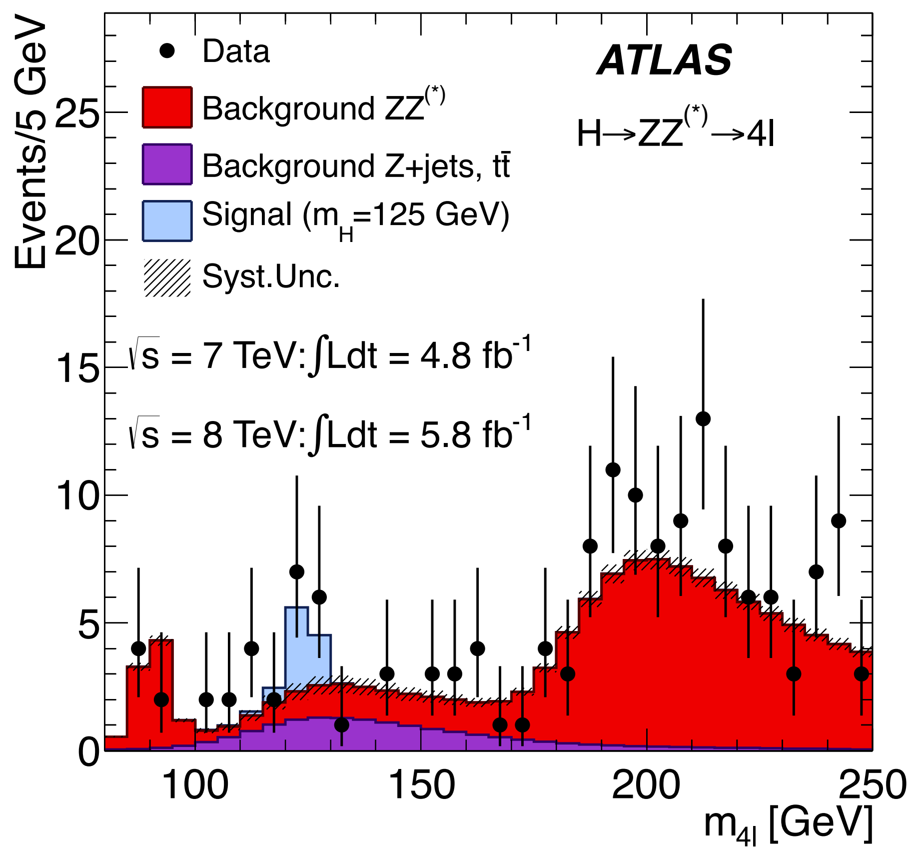
Top quark discovery (1990s)

Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}

Multiple events:
Cut-and-count

CDF Collaboration: arXiv:9503002

Higgs boson discovery (2010s)



Shape information:
Histogram

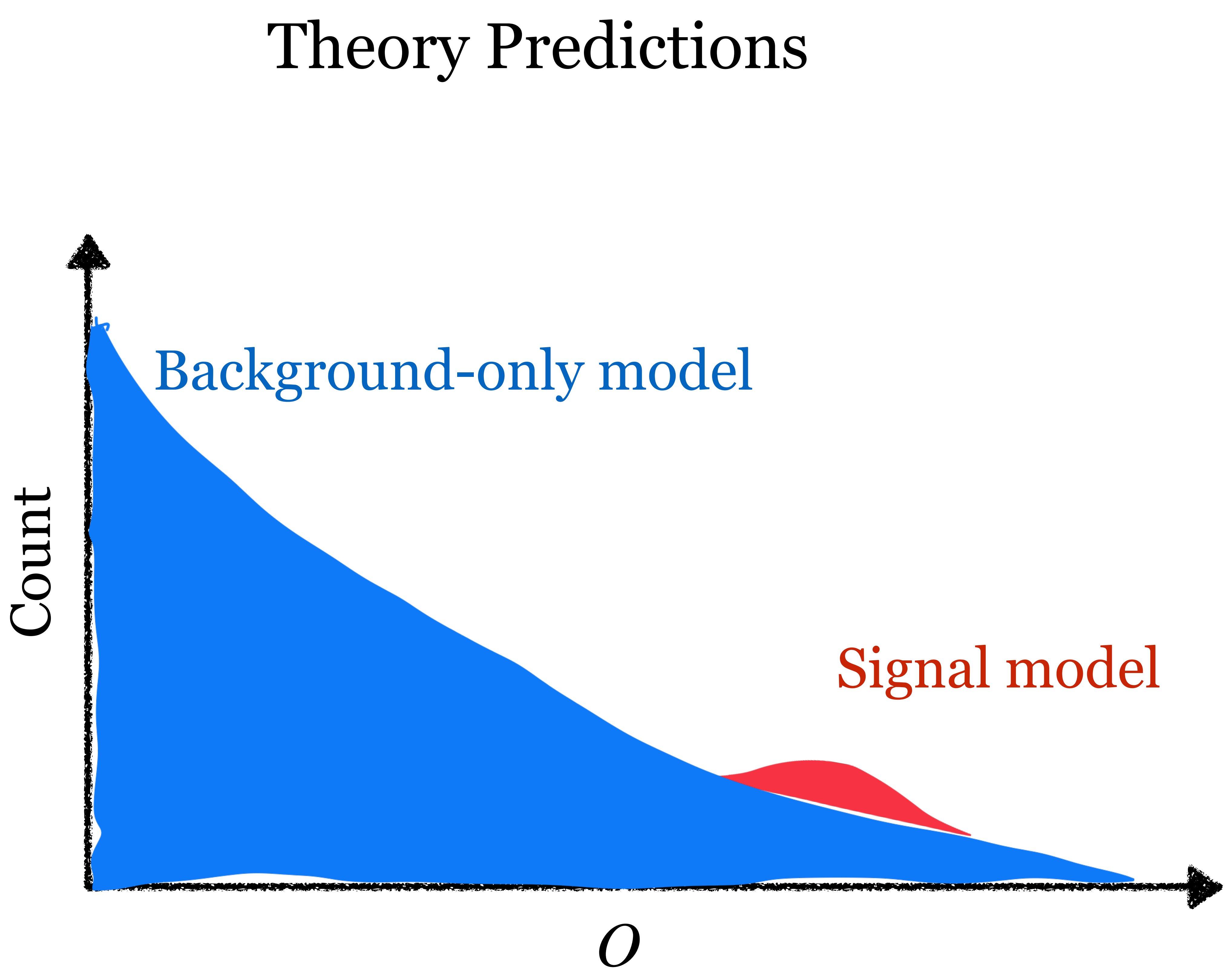
ATLAS Collaboration: arXiv:1207.7214

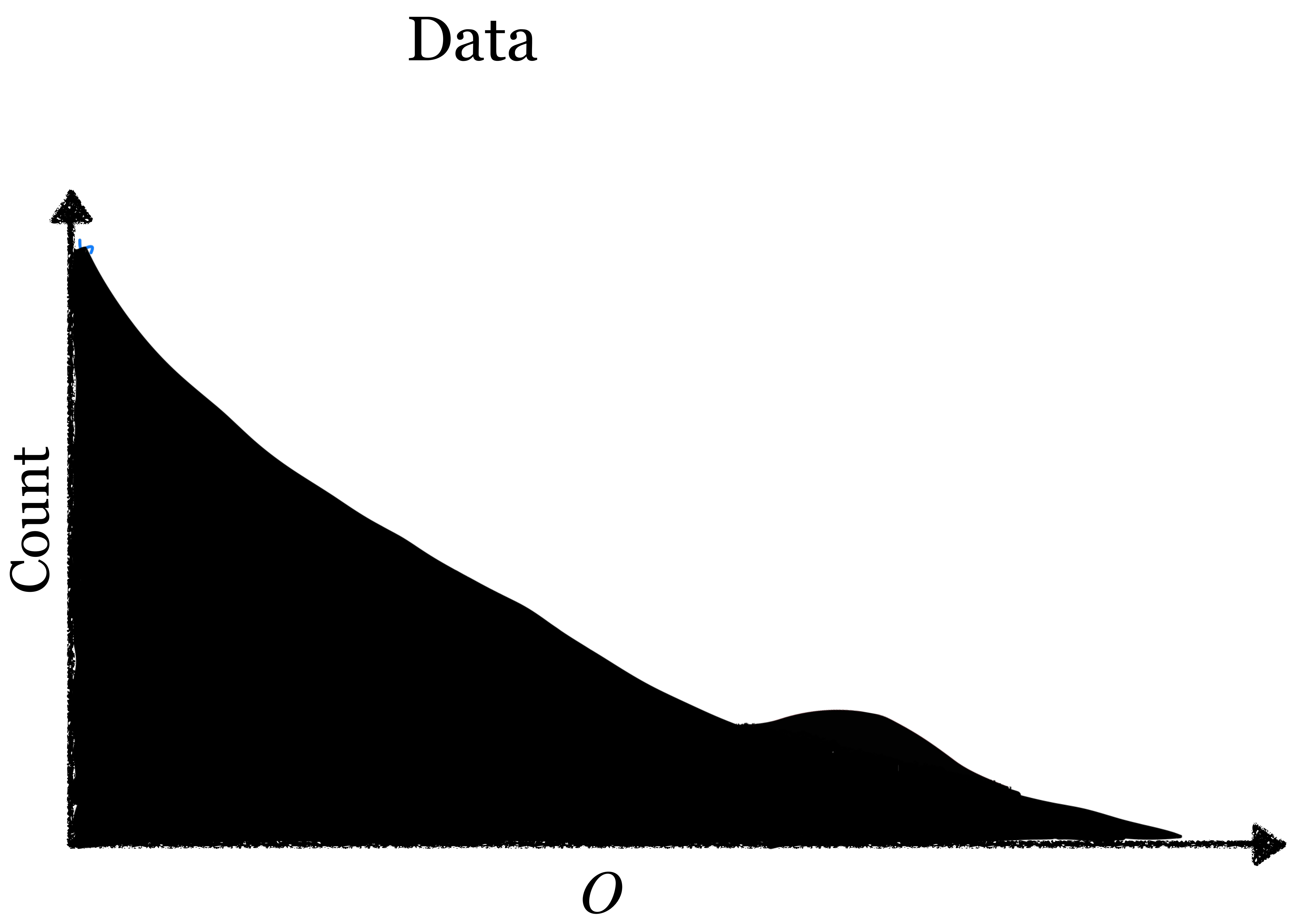
$$p_{bin}(n_{obs} \,|\, n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

$$p(\mathscr{D} \,|\, \mu) = \prod p_{bin}(n_{obs} \,|\, n_{exp})$$

Histogram bins just a bunch of independent counting experiments

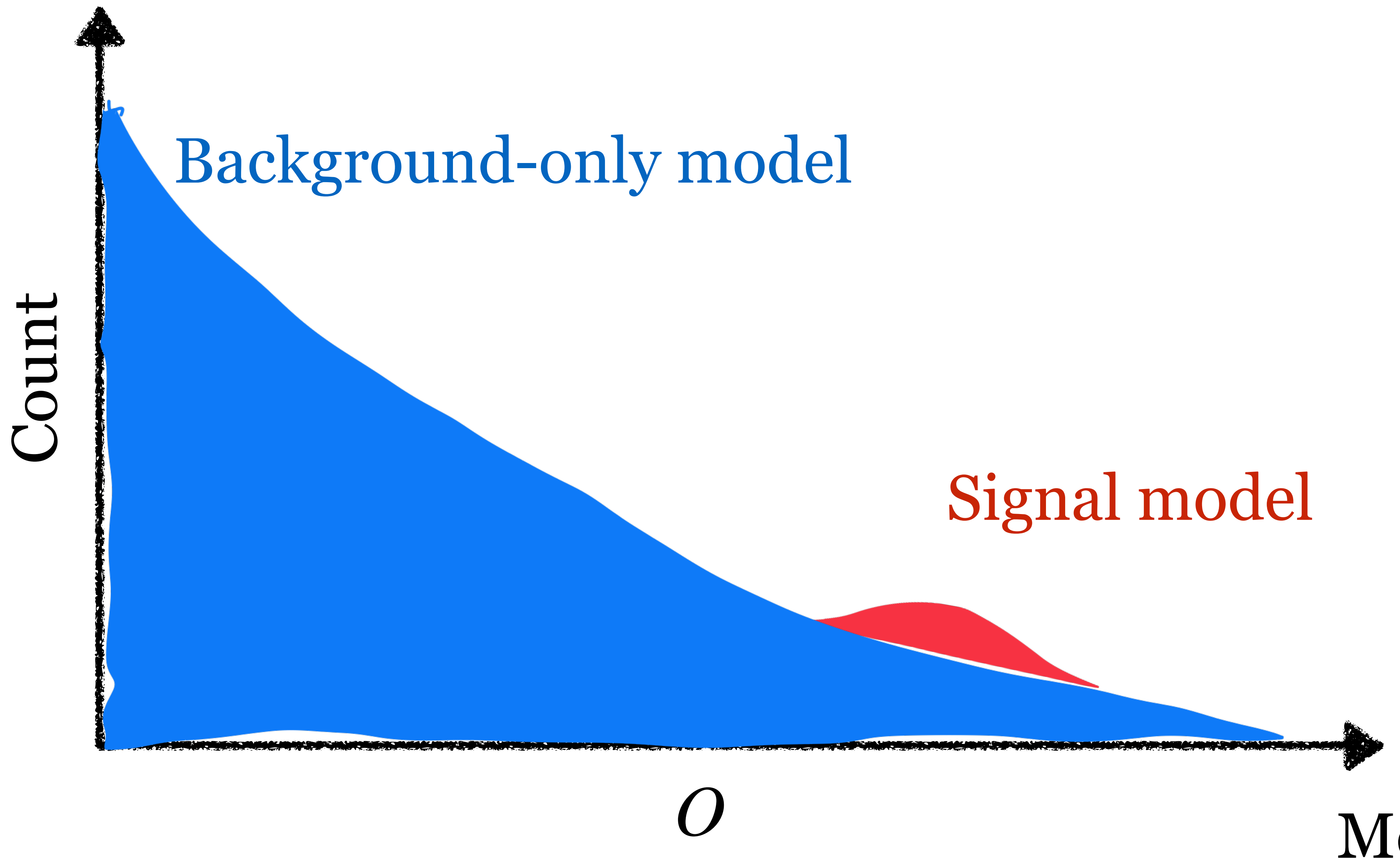
Probability density estimation, with a histogram



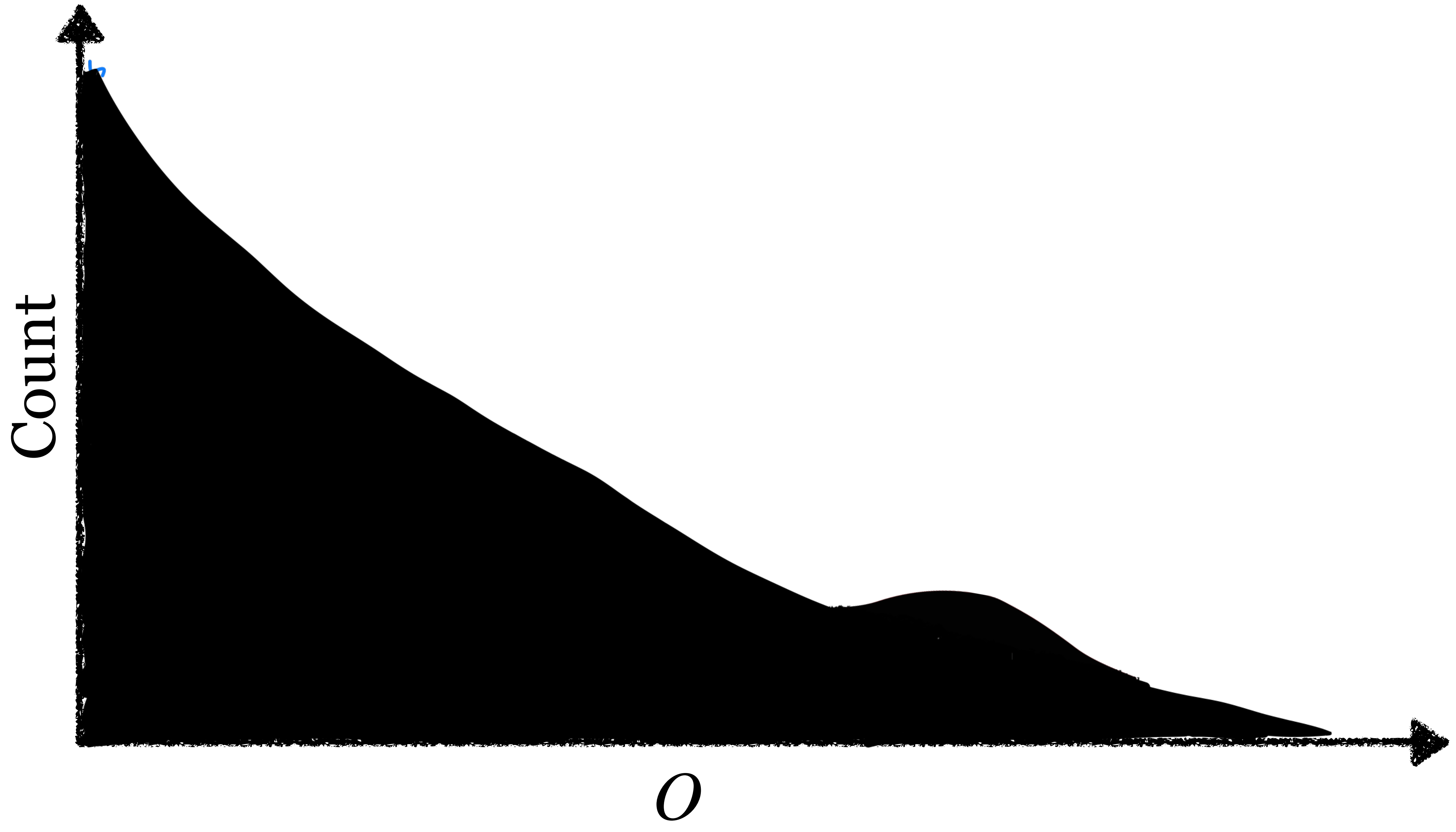


Probability density estimation, with a histogram

Theory Predictions

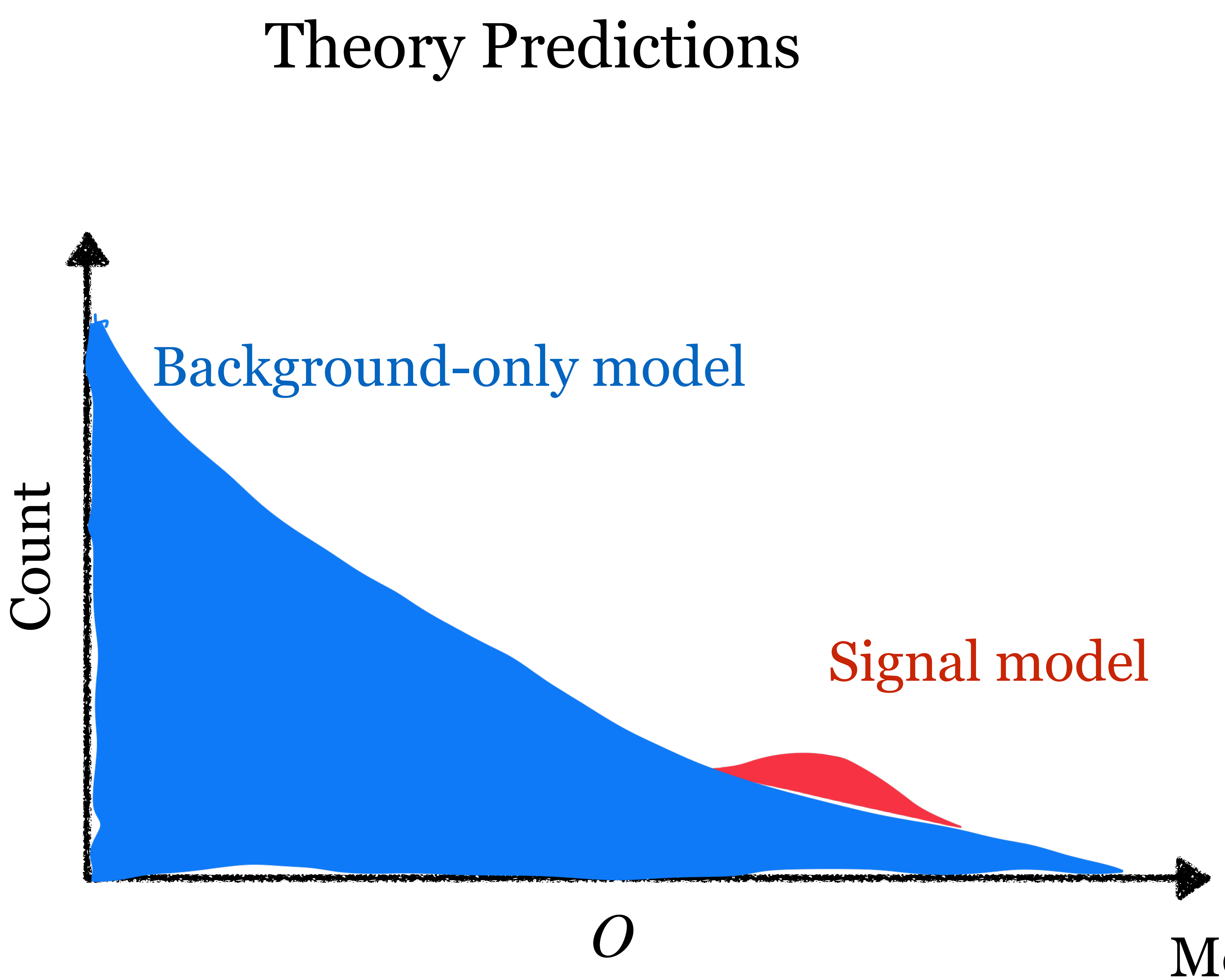


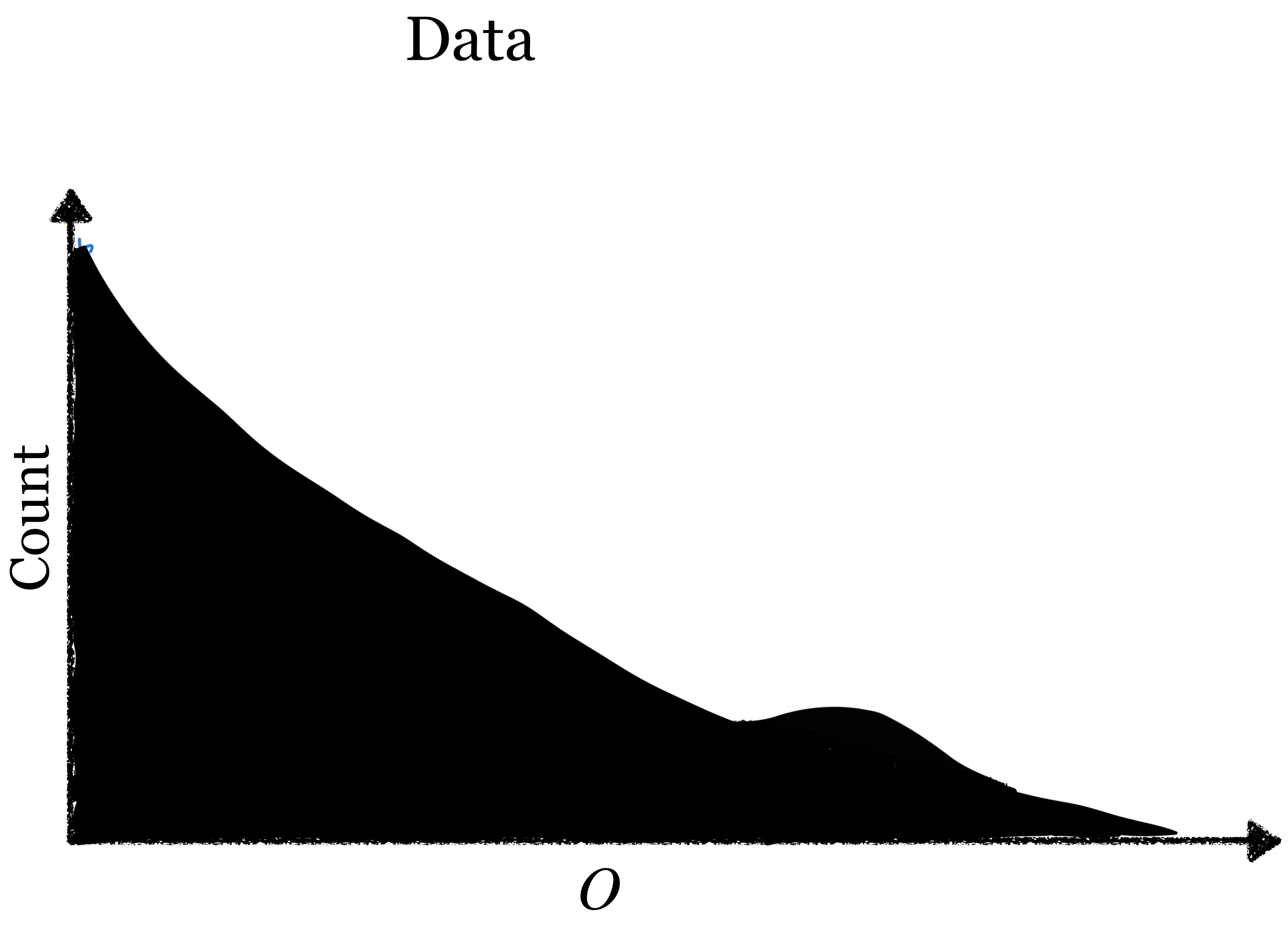
Data



Measure ‘signal strength’ μ

Probability density estimation, with a histogram

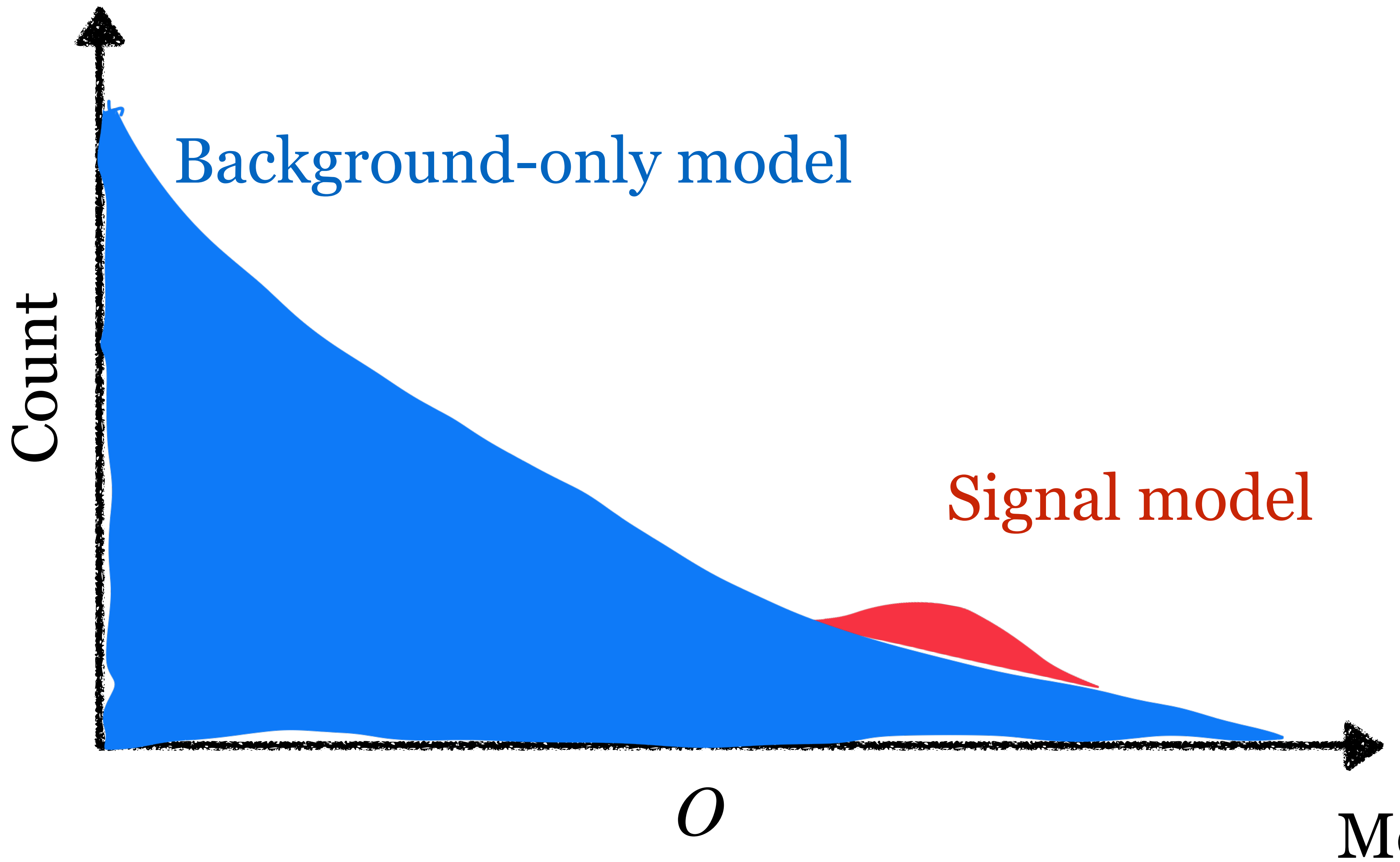




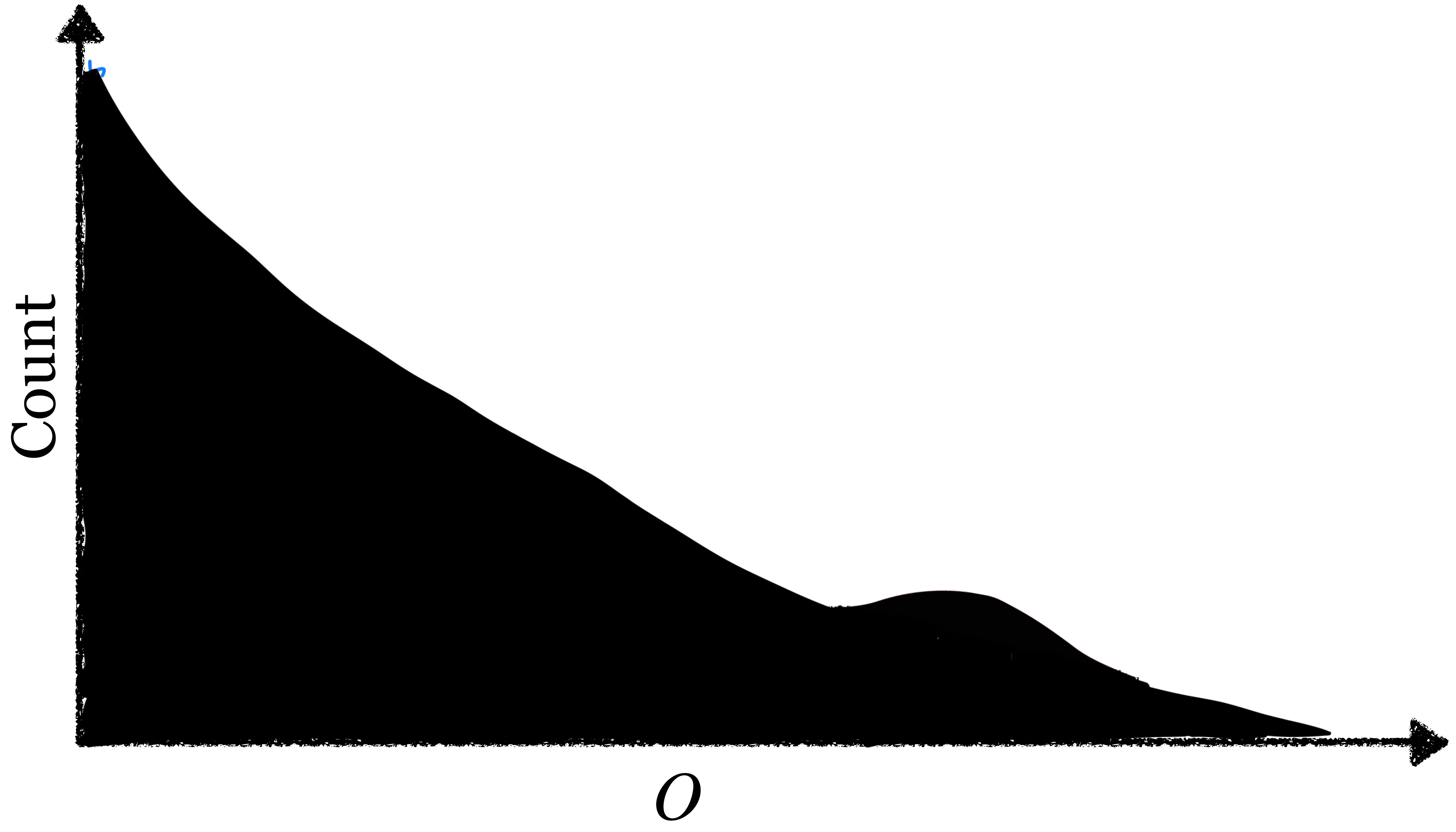
Measure 'signal strength' μ

Probability density estimation, with a histogram

Theory Predictions



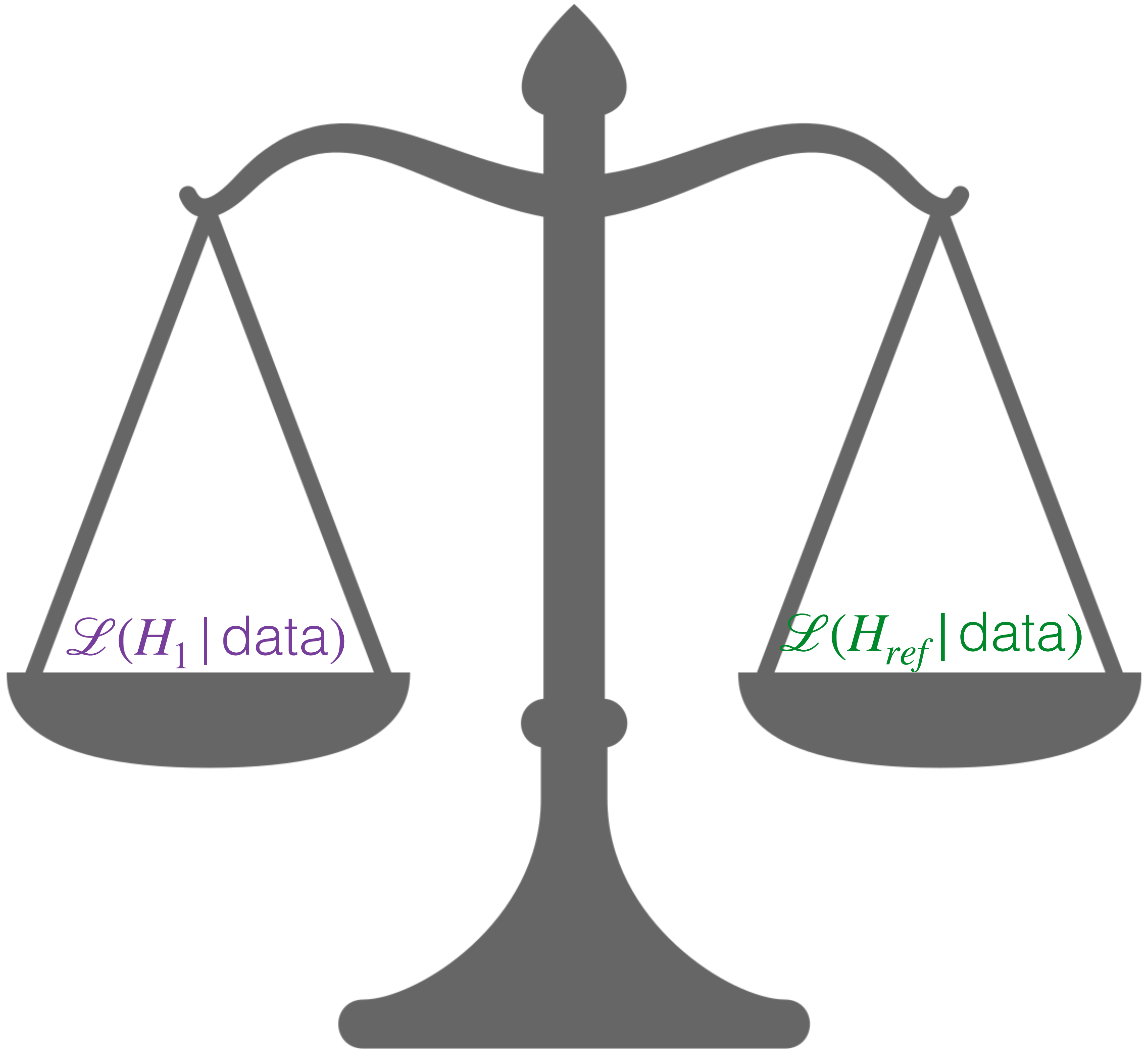
Data



Measure ‘signal strength’ μ

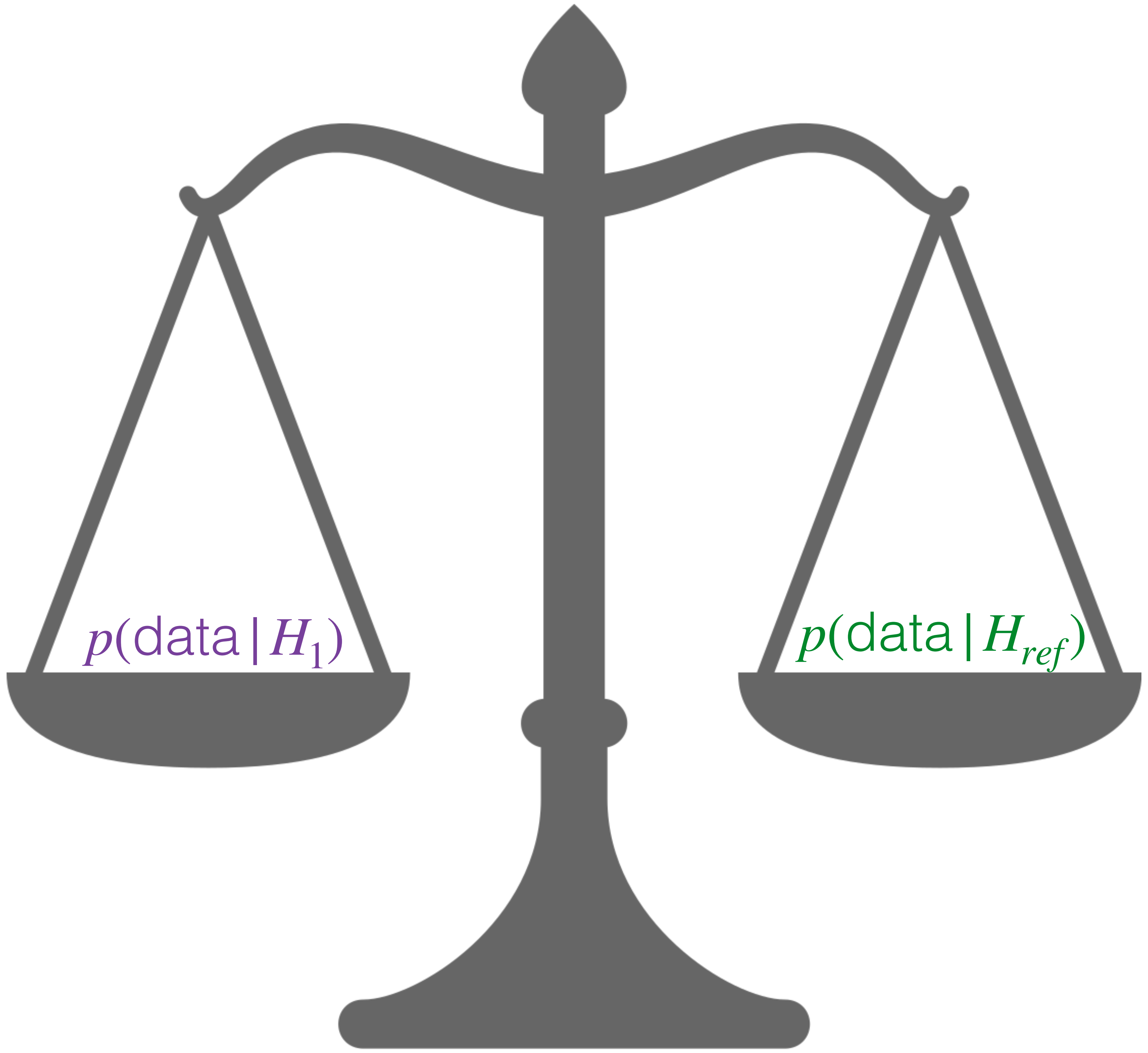
With histograms we can ask “Given the data, what is the likelihood of $\mu = 1$ hypothesis vs $\mu = 2$ hypothesis?”

(Frequentist) Hypothesis tests



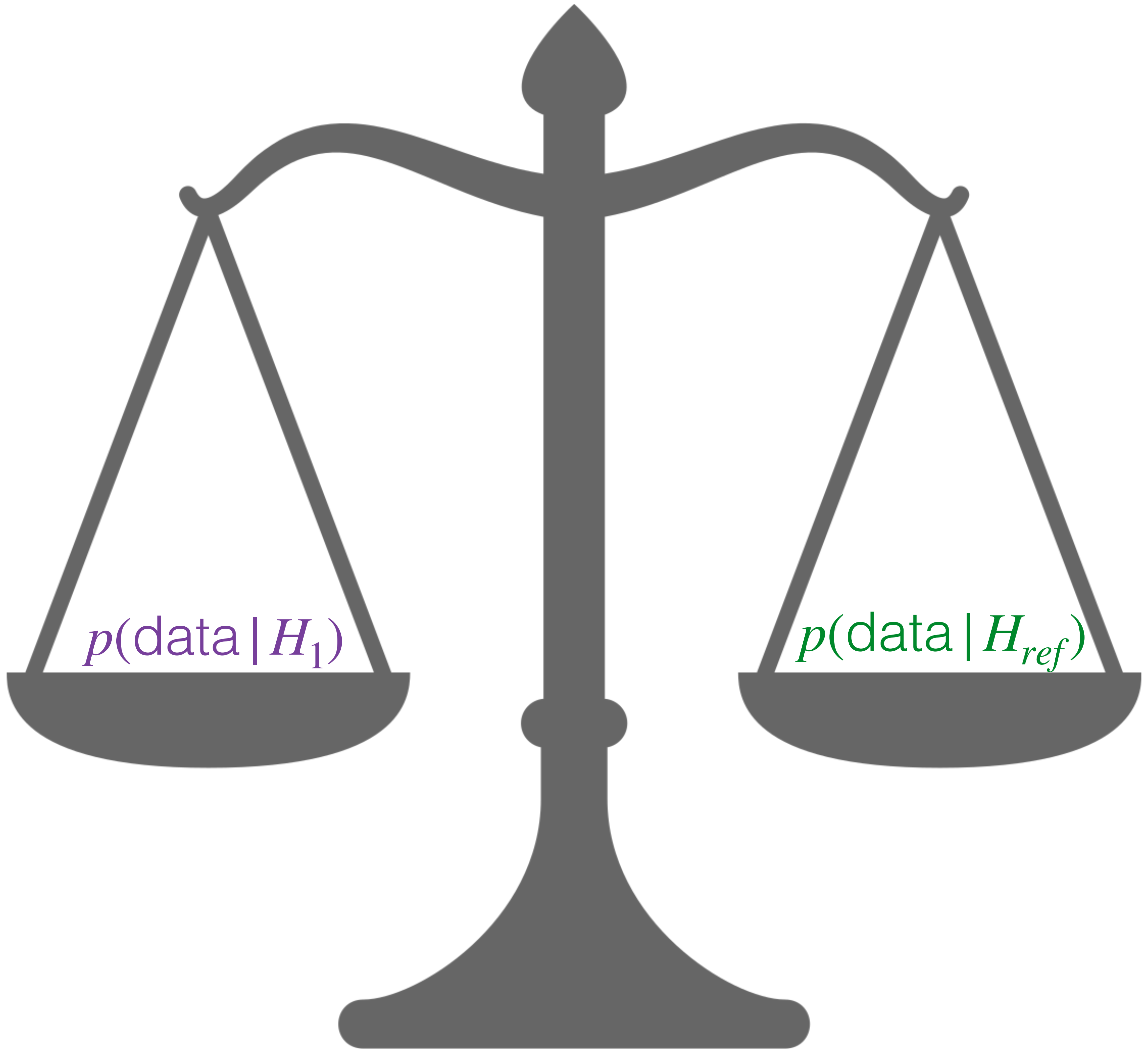
(Frequentist) Hypothesis tests

$$\mathscr{L}(H_1 \mid \text{data}) = p(\text{data} \mid H_1)$$



(Frequentist) Hypothesis tests

$$\mathscr{L}(H_1 | \text{data}) = p(\text{data} | H_1)$$

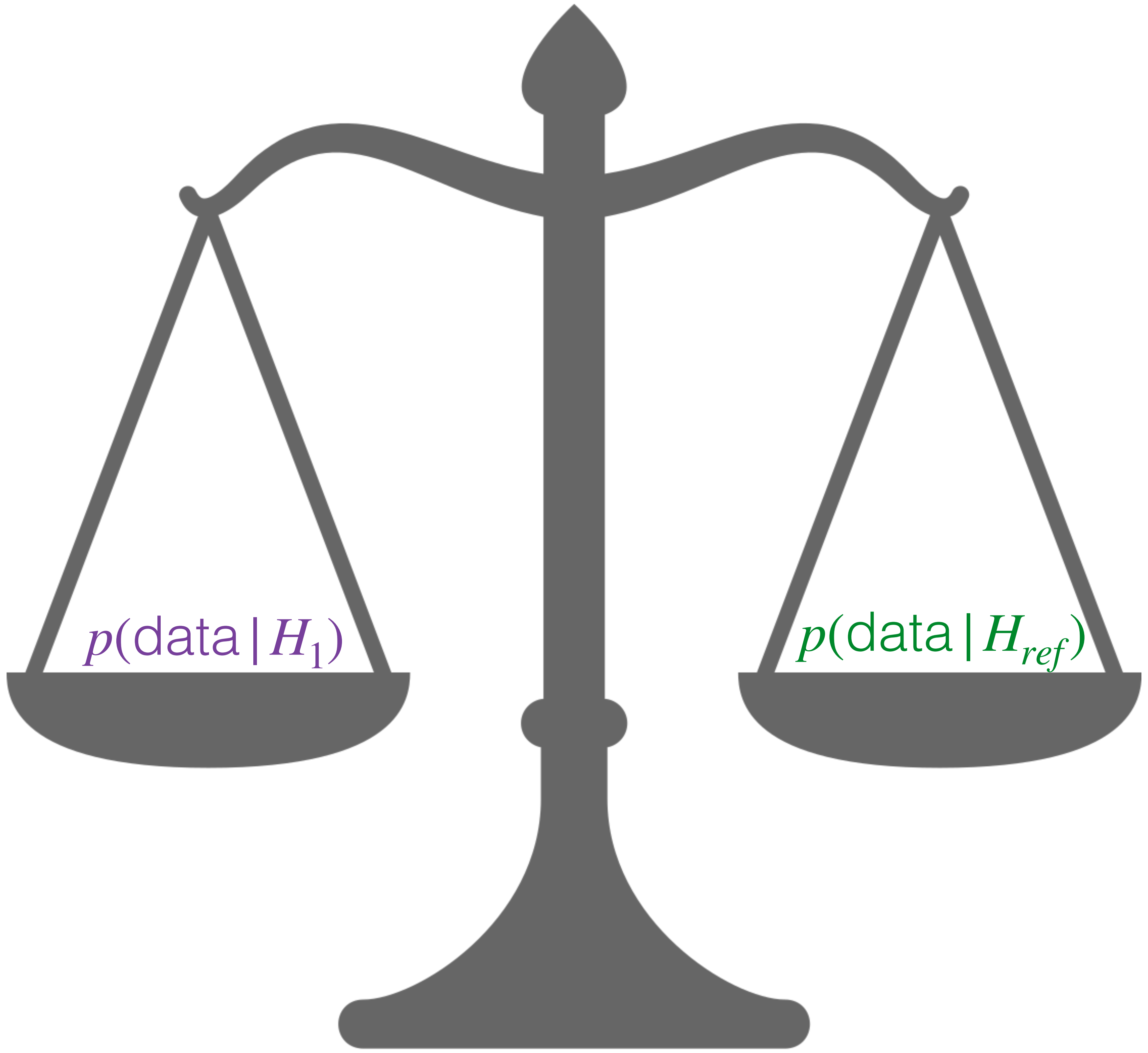


When comparing 2 hypotheses, LR guaranteed to be optimal test by Neyman-Person lemma

(Frequentist) Hypothesis tests

$$\mathscr{L}(H_1 | \text{data}) = p(\text{data} | H_1)$$

$$H_1 = \{\mu_1\} \qquad H_1^1 = \{\mu_1 \, \alpha_1\}$$



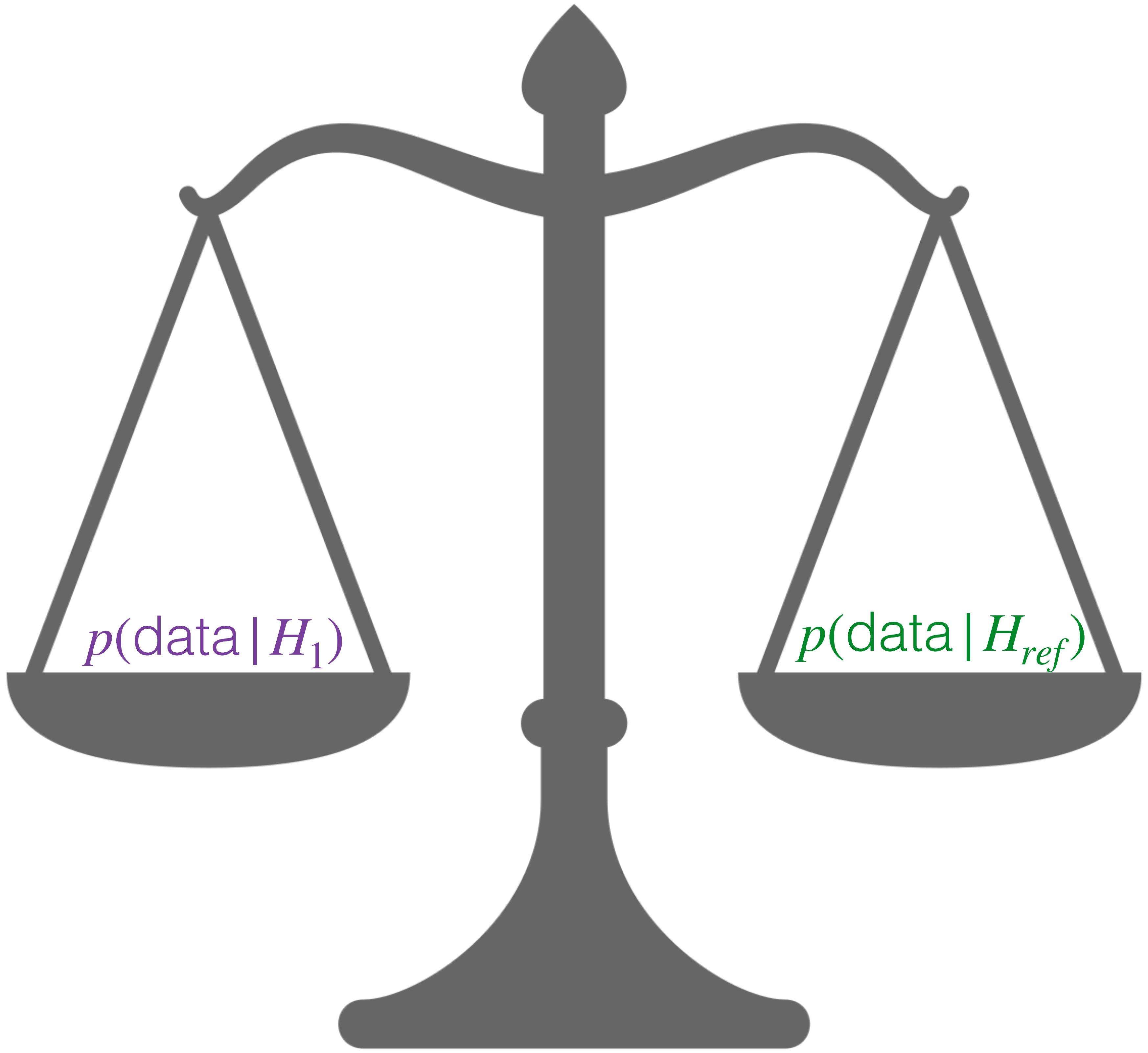
When comparing 2 hypotheses, LR guaranteed to be optimal test by Neyman-Person lemma

μ: Parameter of interest, you want to measure this

α: Nuisance parameters, you need to measure this to avoid biases. Systematic uncertainties like jet energy scale

(Frequentist) Hypothesis tests

$$\mathscr{L}(H_1 \,|\, \text{data}) = p(\text{data} \,|\, H_1)$$



When comparing 2 hypotheses, LR guaranteed to be optimal test by Neyman-Person lemma

$H_1 = \{\mu_1\}$ $\quad H_1^1 = \{\mu_1 \alpha_1\}$

Likelihood $\mathscr{L}(\mu, \alpha \,|\, \mathscr{D})$: $\quad p(\mathscr{D} \,|\, \mu, \alpha)$

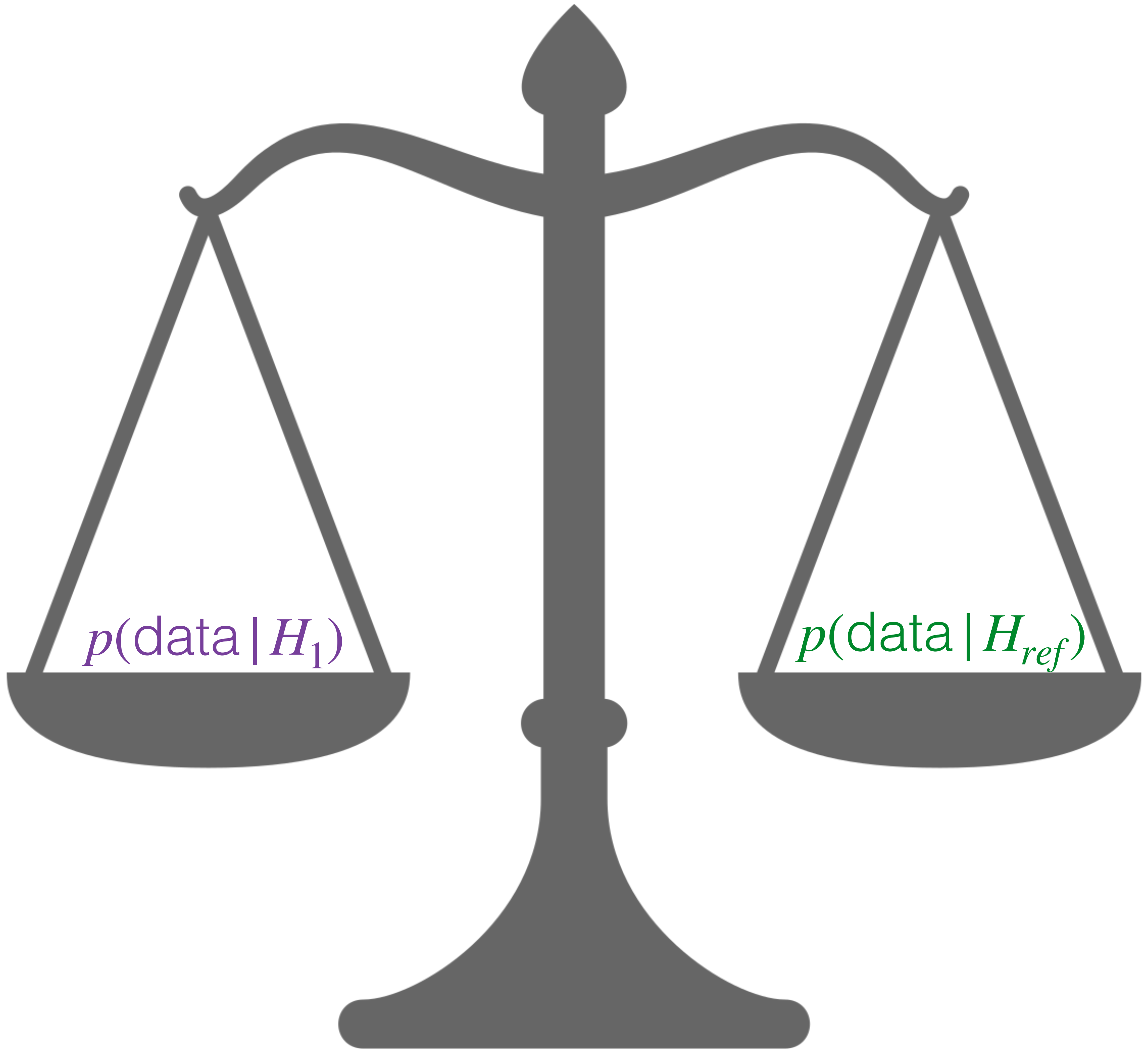
μ: Parameter of interest, you want to measure this

α: Nuisance parameters, you need to measure this to avoid biases. Systematic uncertainties like jet energy scale

$\mathscr{D}$ now is a 1D summary of the high-dimensional data

(Frequentist) Hypothesis tests

$$\mathscr{L}(H_1 \,|\, \text{data}) = p(\text{data} \,|\, H_1)$$



When comparing 2 hypotheses, LR guaranteed to be optimal test by Neyman-Person lemma

$H_1 = \{\mu_1\}$ $\quad$ $H_1^1 = \{\mu_1 \, \alpha_1\}$

Likelihood $\mathscr{L}(\mu, \alpha \,|\, \mathscr{D})$: $\quad p(\mathscr{D} \,|\, \mu, \alpha)$

Profile likelihood ratio: $\quad \dfrac{p(\mathscr{D} \,|\, \mu, \hat{\hat{\alpha}})}{p(\mathscr{D} \,|\, \hat{\mu}, \hat{\alpha})}$

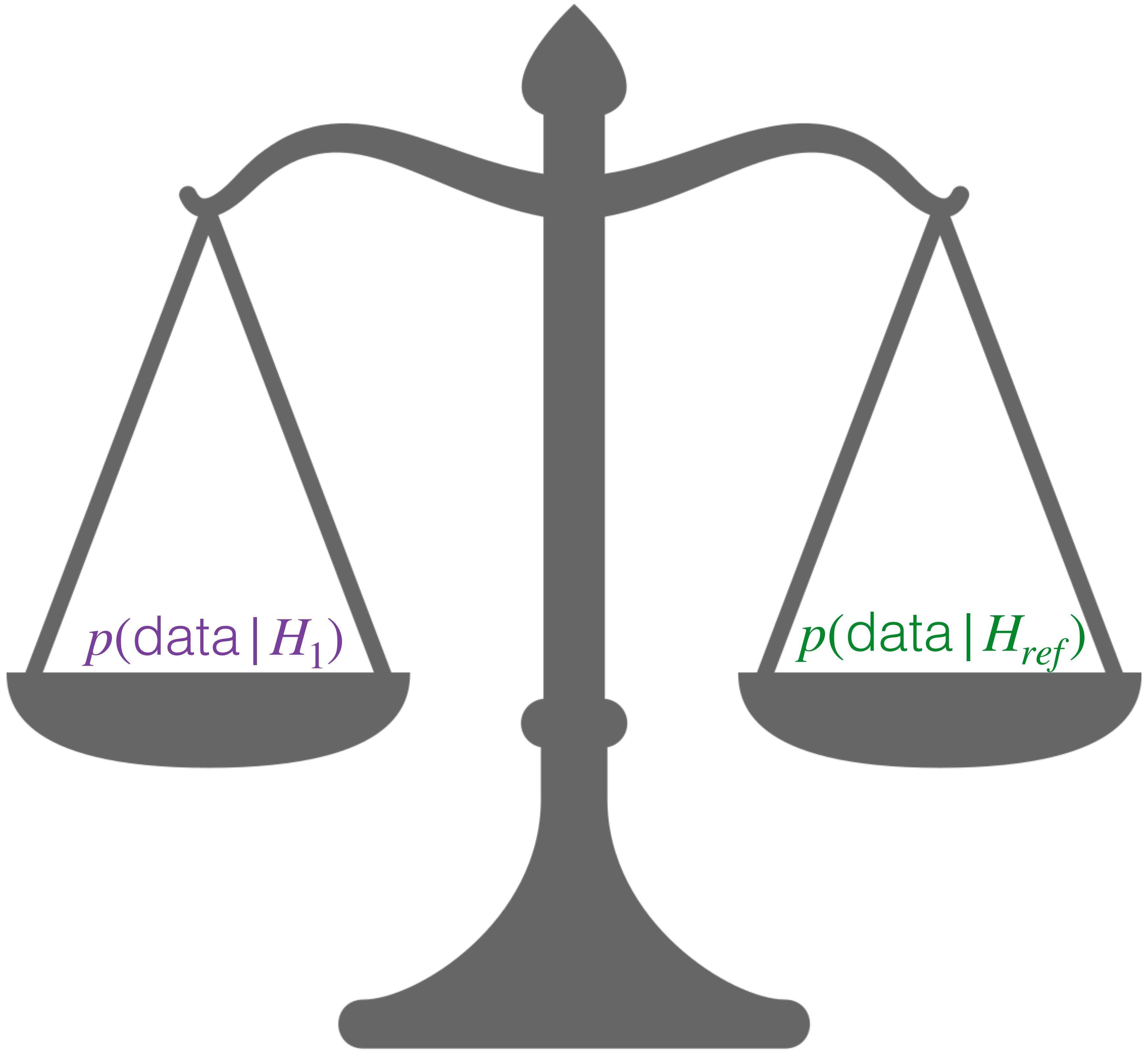
μ: Parameter of interest, you want to measure this

α: Nuisance parameters, you need to measure this to avoid biases. Systematic uncertainties like jet energy scale

$\mathscr{D}$ now is a 1D summary of the high-dimensional data

(Frequentist) Hypothesis tests

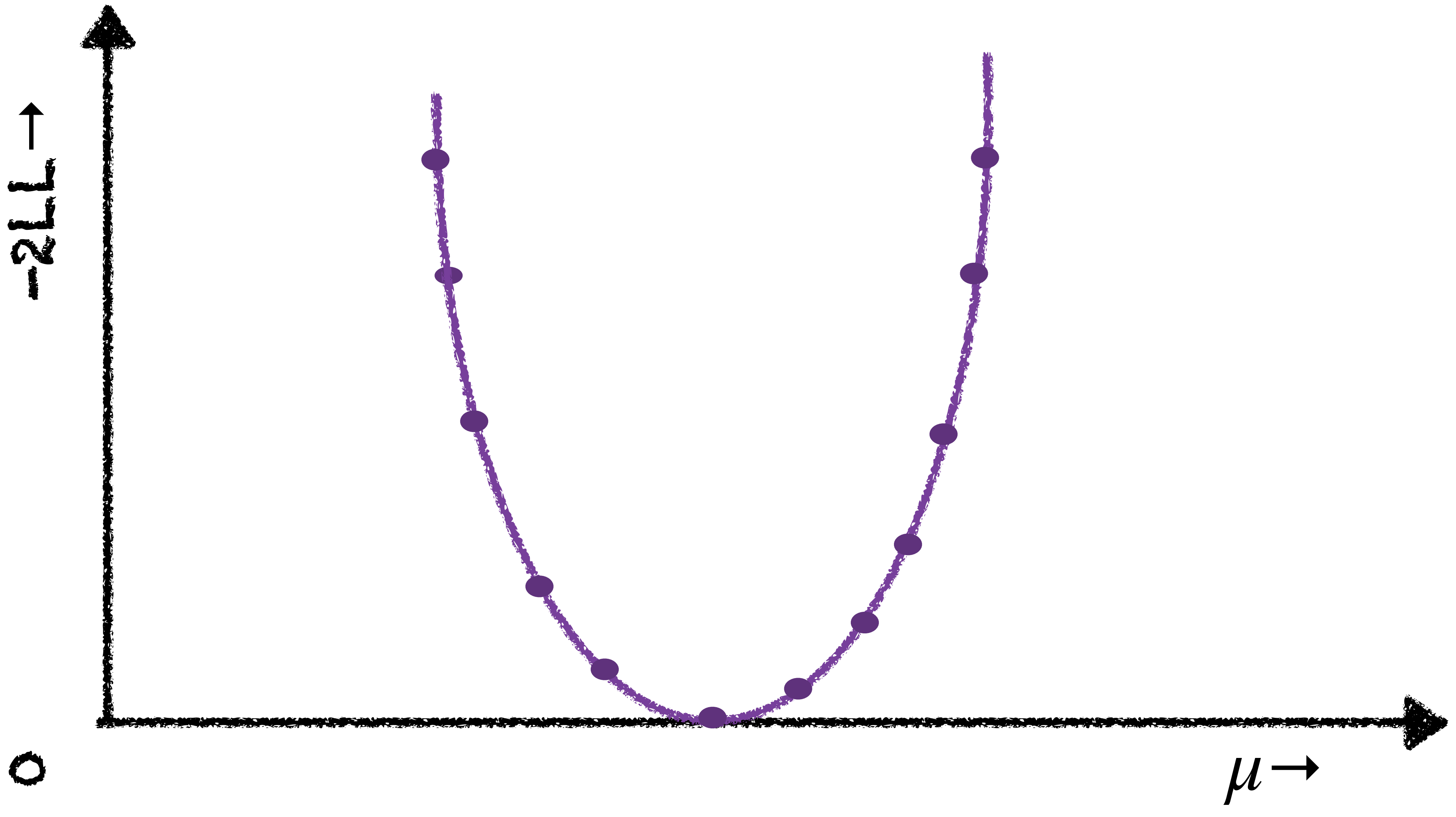
$$\mathscr{L}(H_1 \,|\, \text{data}) = p(\text{data} \,|\, H_1)$$



When comparing 2 hypotheses, LR guaranteed to be optimal test by Neyman-Person lemma

Invert the test to get confidence intervals

$$-2 \log \frac{p(\mathscr{D} \,|\, \mu, \hat{\hat{\alpha}})}{p(\mathscr{D} \,|\, \hat{\mu}, \hat{\alpha})}$$

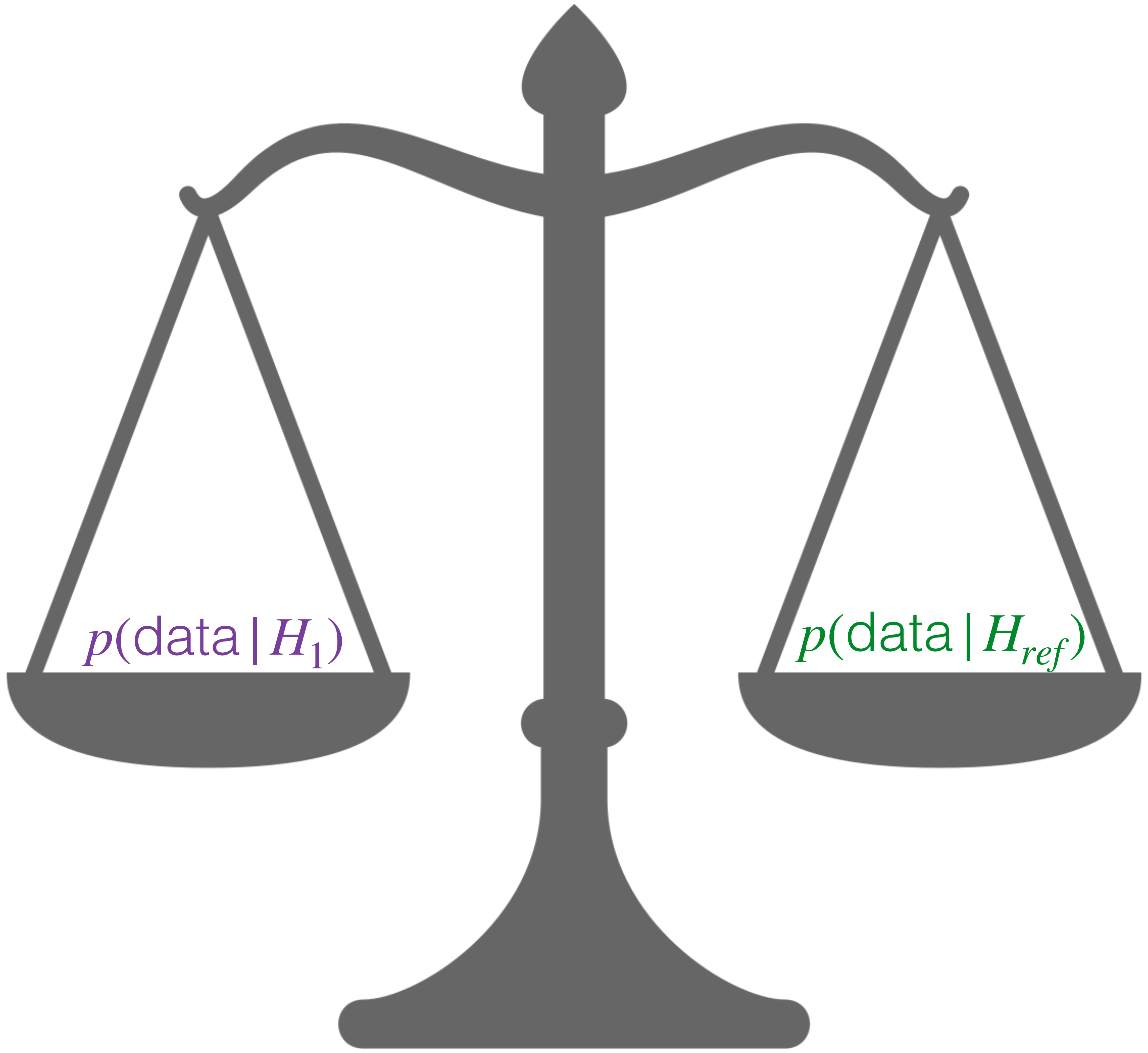


Parameter estimation (infinite hypotheses)

$\mathscr{D}$ now is a 1D summary of the high-dimensional data

(Frequentist) Hypothesis tests

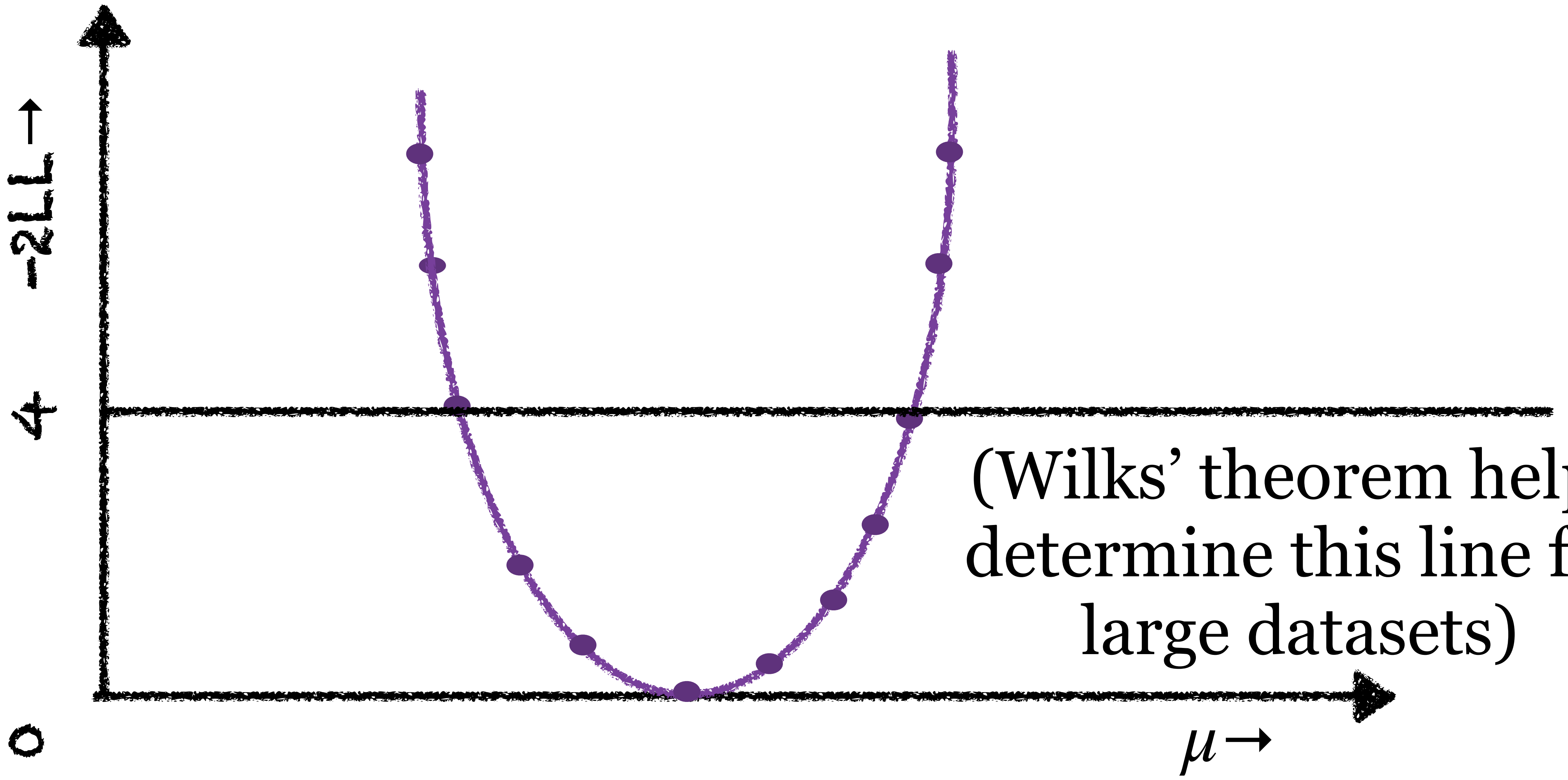
$$\mathscr{L}(H_1 \mid \text{data}) = p(\text{data} \mid H_1)$$



When comparing 2 hypotheses, LR guaranteed to be optimal test by Neyman-Person lemma

Invert the test to get confidence intervals

$$-2 \log \frac{p(\mathscr{D} \mid \mu, \hat{\hat{\alpha}})}{p(\mathscr{D} \mid \hat{\mu}, \hat{\alpha})}$$

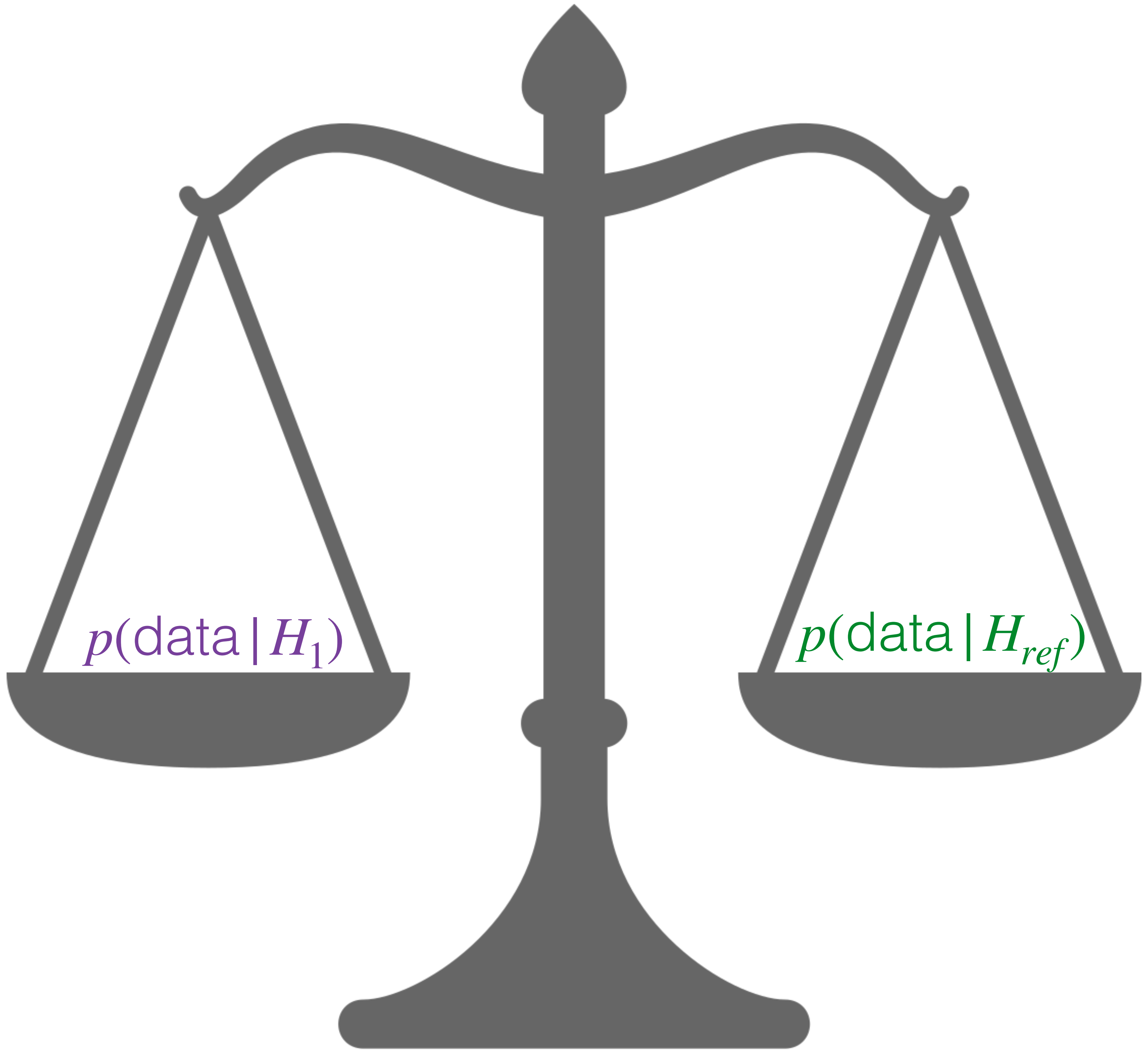


(Wilks' theorem helps determine this line for large datasets)

Parameter estimation (infinite hypotheses)

(Frequentist) Hypothesis tests

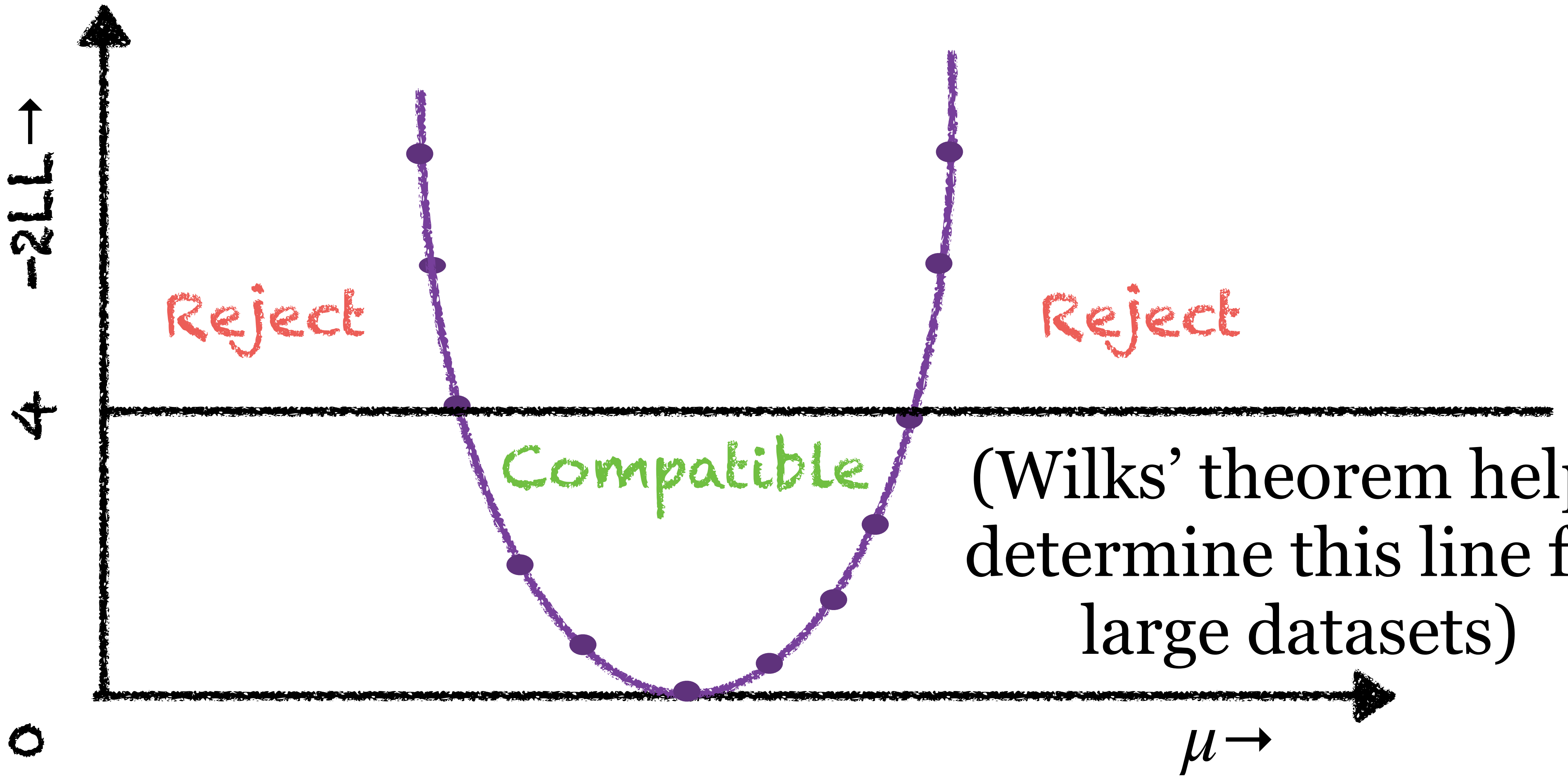
$$\mathscr{L}(H_1 \,|\, \text{data}) = p(\text{data} \,|\, H_1)$$



When comparing 2 hypotheses, LR guaranteed to be optimal test by Neyman-Person lemma

Invert the test to get confidence intervals

$$-2 \log \frac{p(\mathscr{D} \,|\, \mu, \hat{\hat{\alpha}})}{p(\mathscr{D} \,|\, \hat{\mu}, \hat{\alpha})}$$

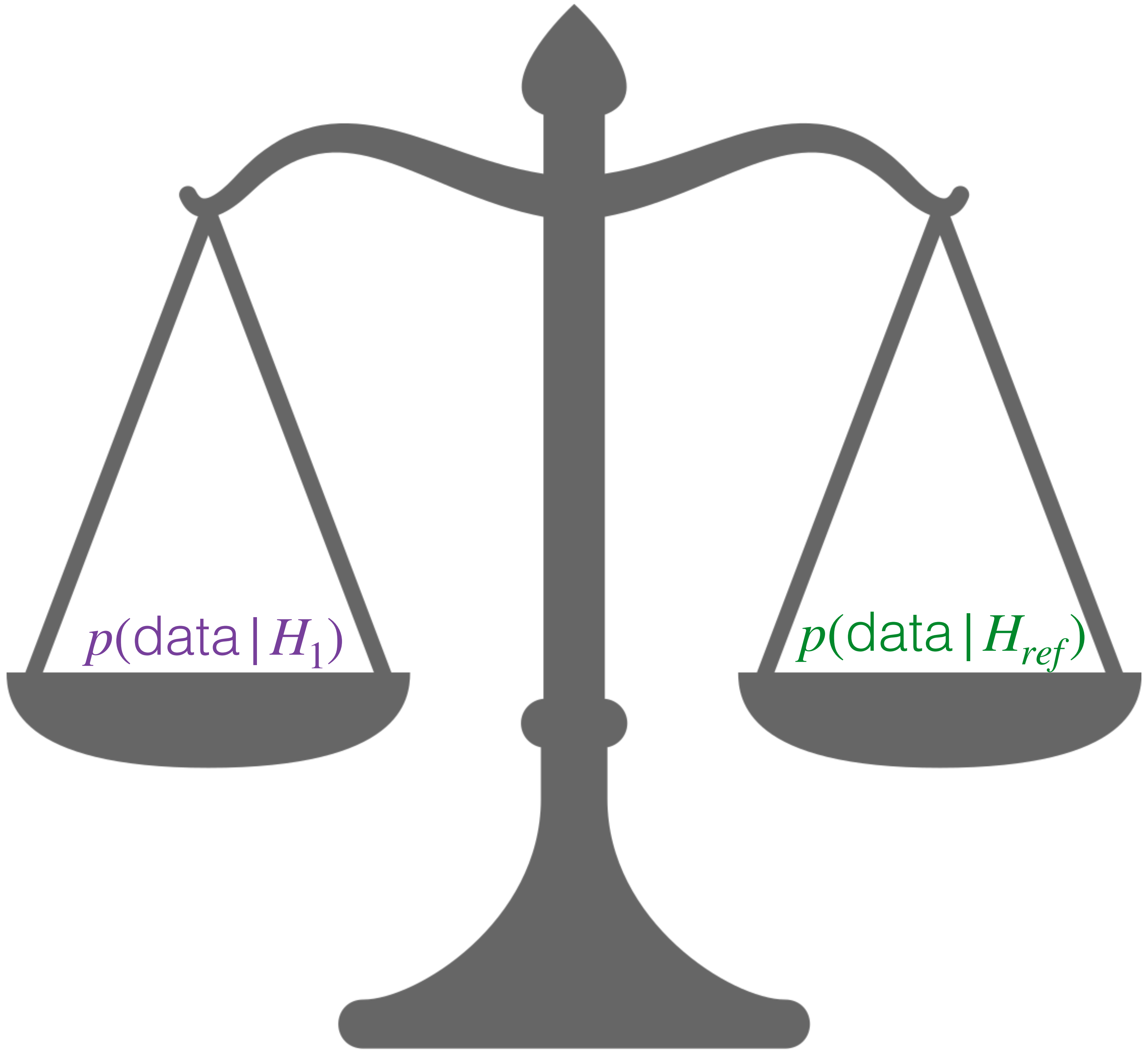


(Wilks' theorem helps determine this line for large datasets)

Parameter estimation (infinite hypotheses)

(Frequentist) Hypothesis tests

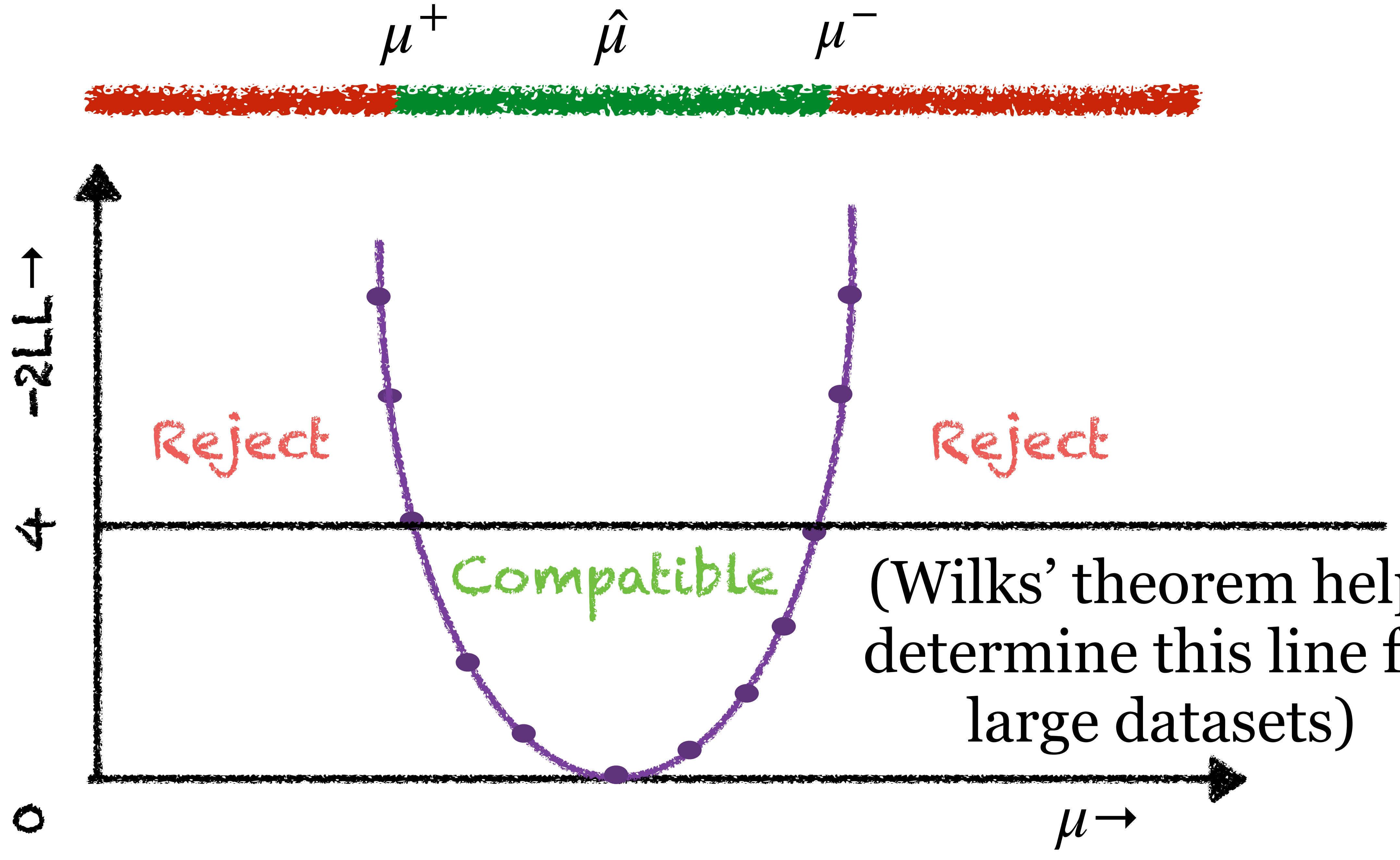
$$\mathscr{L}(H_1 \mid \text{data}) = p(\text{data} \mid H_1)$$



When comparing 2 hypotheses, LR guaranteed to be optimal test by Neyman-Person lemma

Invert the test to get confidence intervals

$$-2 \log \frac{p(\mathscr{D} \mid \mu, \hat{\hat{\alpha}})}{p(\mathscr{D} \mid \hat{\mu}, \hat{\alpha})}$$



(Wilks' theorem helps determine this line for large datasets)

Parameter estimation (infinite hypotheses)

$\mathscr{D}$ now is a 1D summary of the high-dimensional data

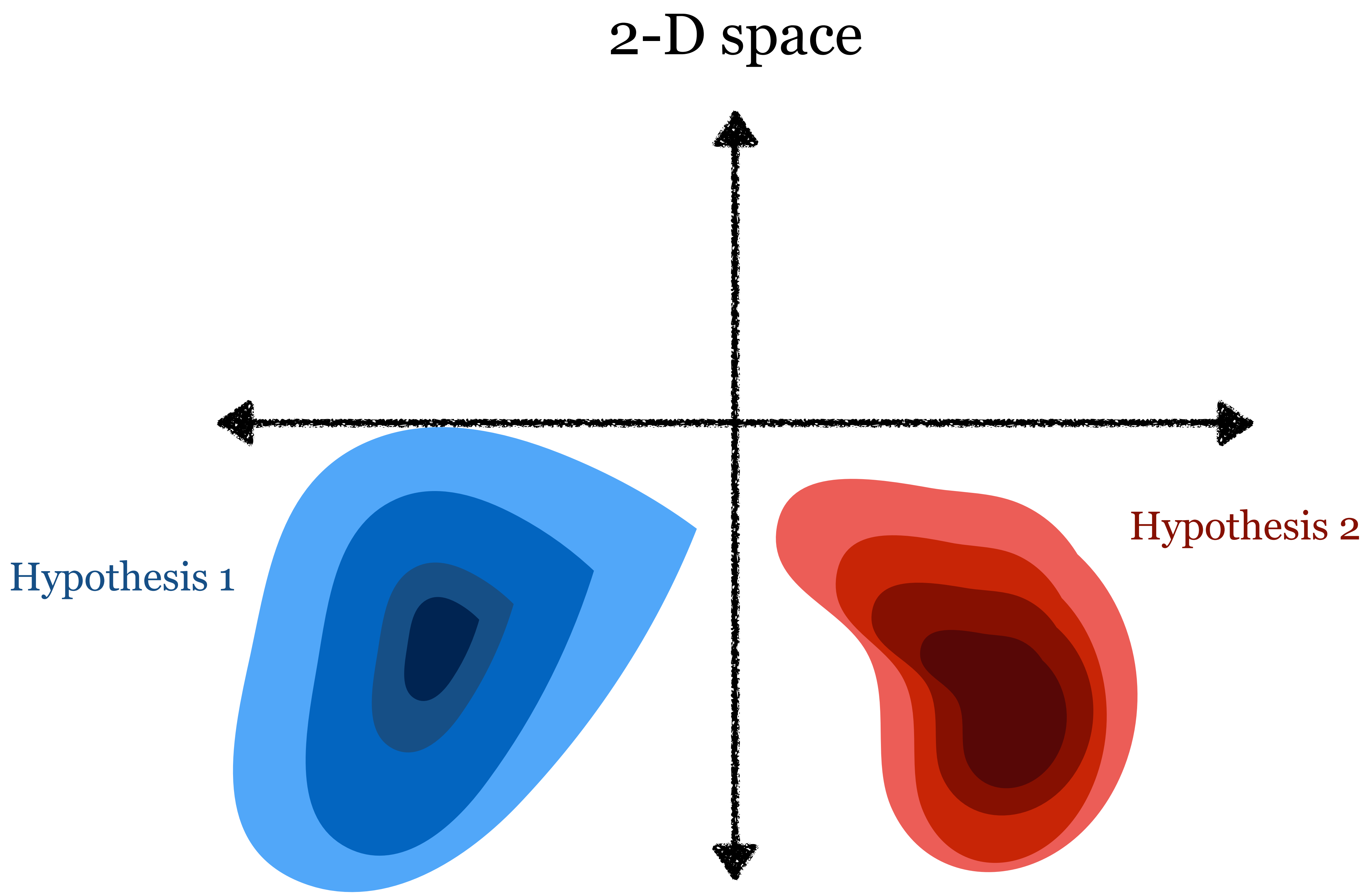
$\mathscr{D}$ now is a 1D summary of the high-dimensional data

Foundational NSBI work: Cranmer et al: arXiv:1506.02169,
https://simulation-based-inference.org/

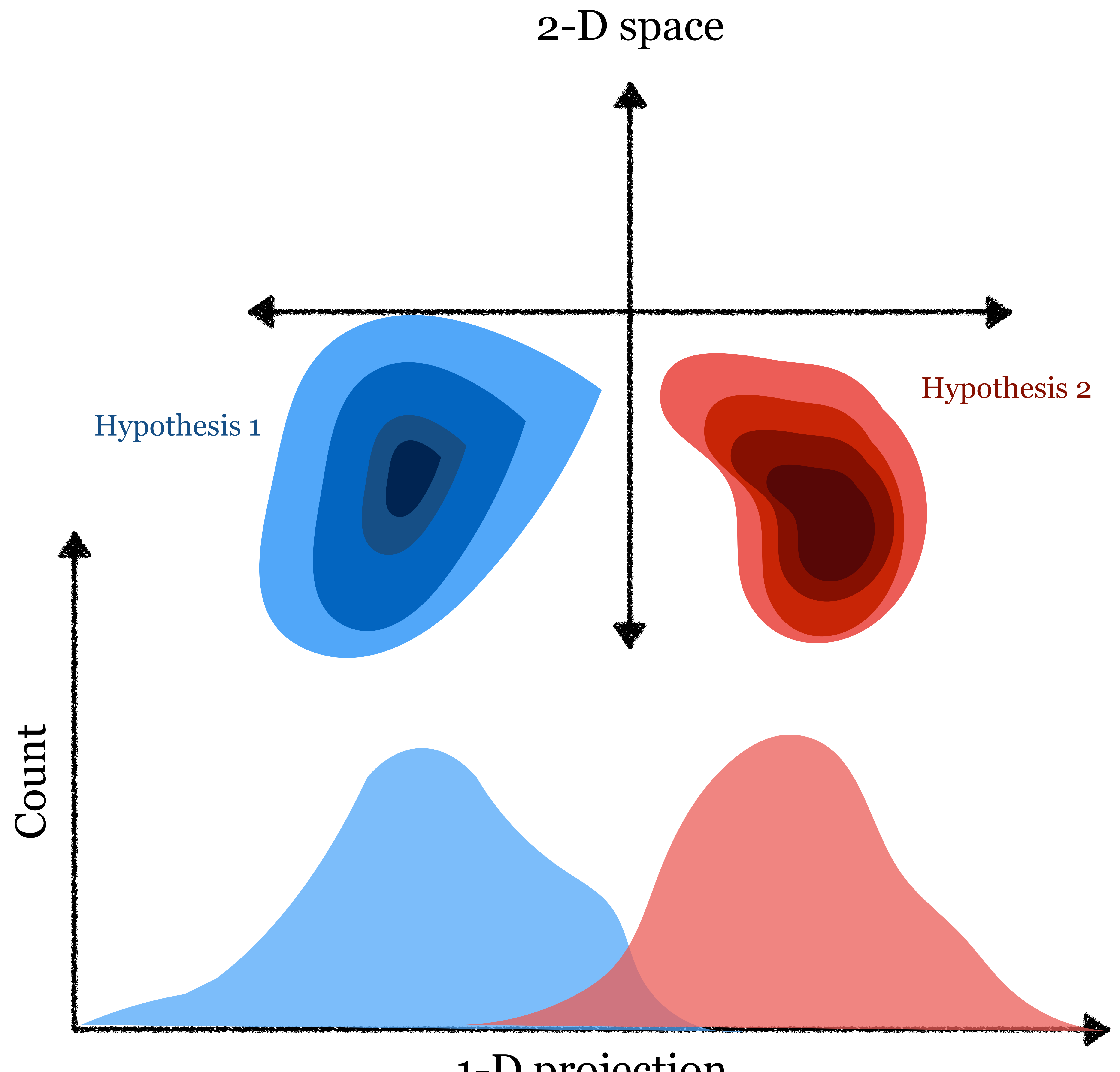
Can we always find a sufficient 1-D summary?

Foundational NSBI work: Cranmer et al: arXiv:1506.02169,
https://simulation-based-inference.org/

Can we always find a sufficient 1-D summary?

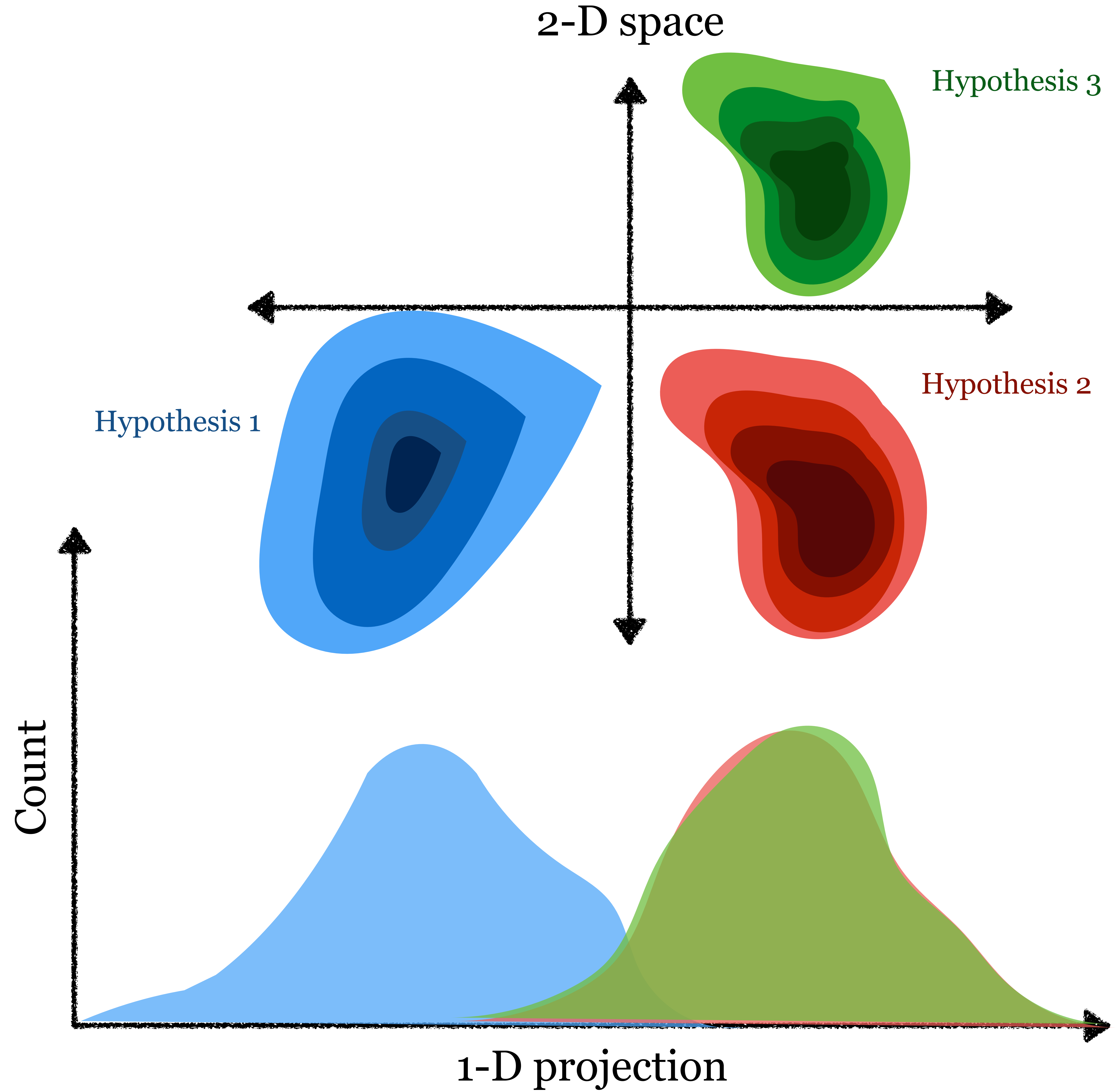


Can we always find a sufficient 1-D summary?



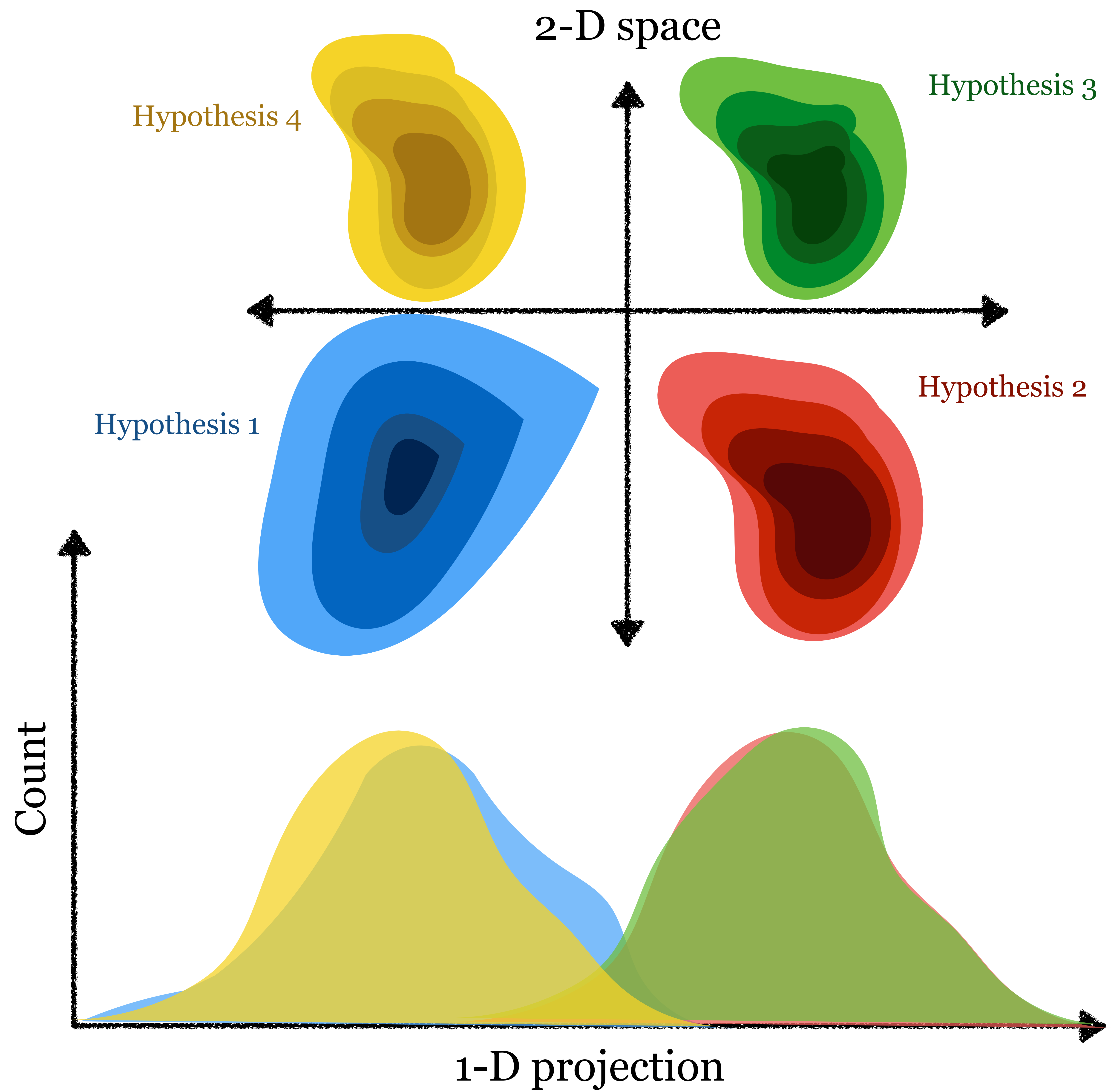
Foundational NSBI work: Cranmer et al: arXiv:1506.02169,
https://simulation-based-inference.org/

Can we always find a sufficient 1-D summary?



Foundational NSBI work: Cranmer et al: arXiv:1506.02169,
https://simulation-based-inference.org/

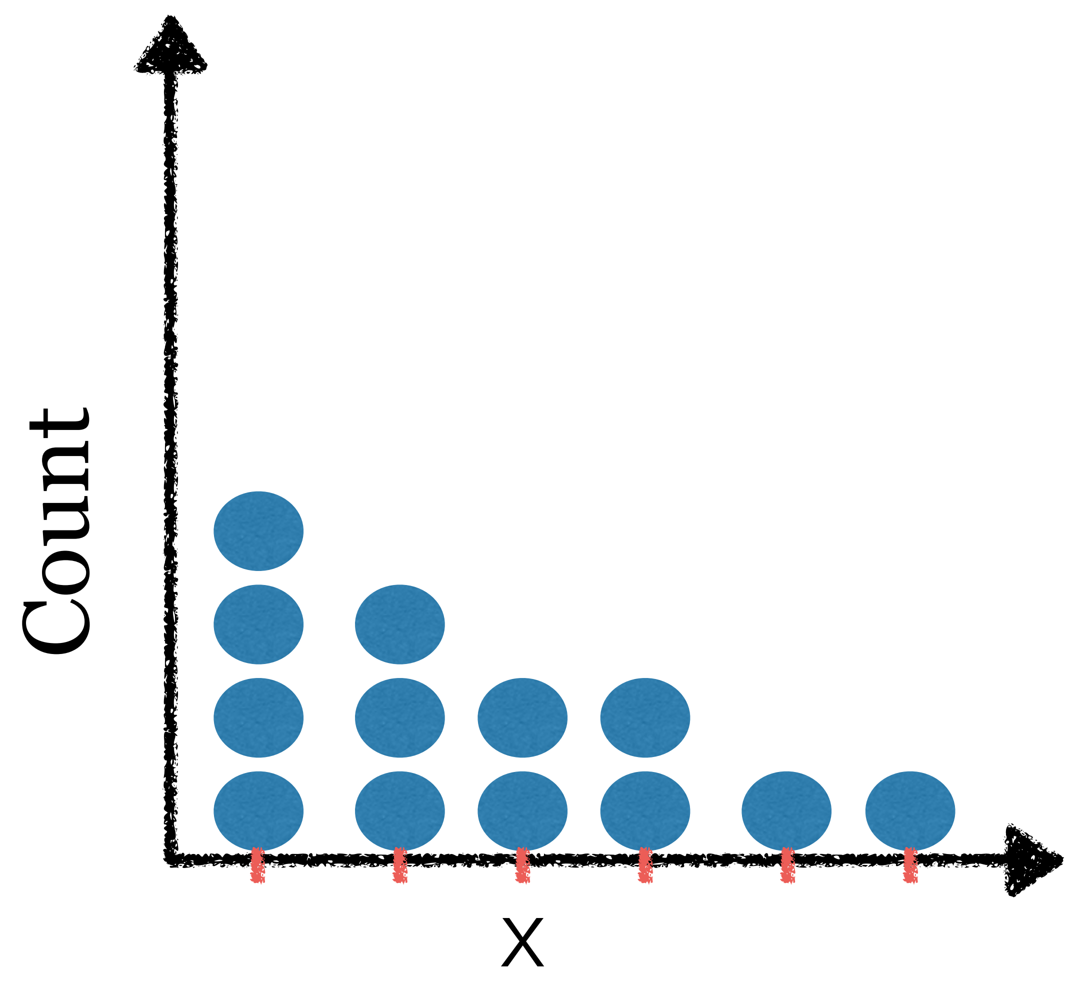
Can we always find a sufficient 1-D summary?



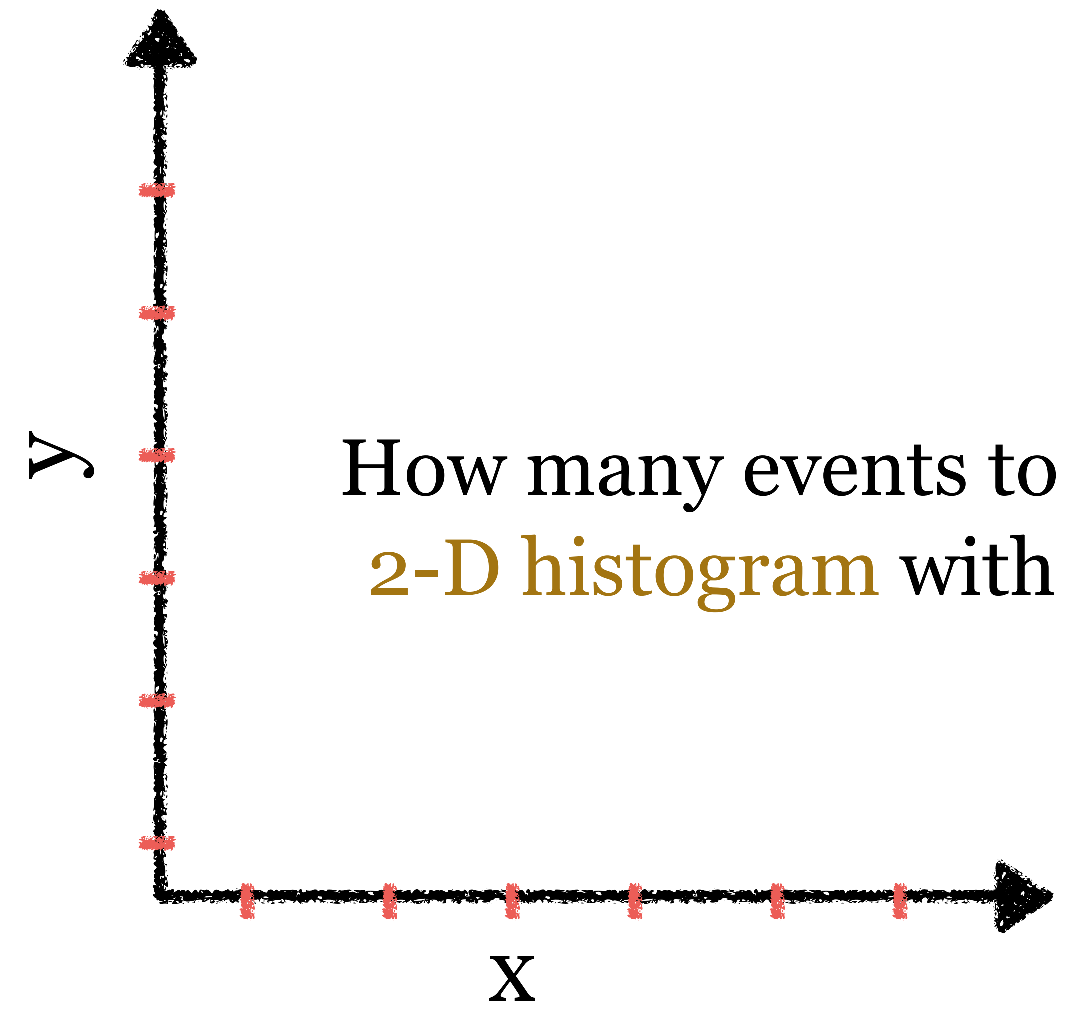
- Clearly separable in 2-D
- A 1-D sufficient summary statistic does not exist for the most important particle physics measurements
  - Higgs Width: hal-02971995(p172): Ghosh et al
  - Systematic uncertainties: PRD 104.056026: Ghosh et a
  - Effective Field Theories: PRD 98.052004: Brehmer et al,
  (Also see talk by Eddie McGrady at CHEP)

Need to perform inference directly in high dim

But probability density estimation in higher dimensions is hard...



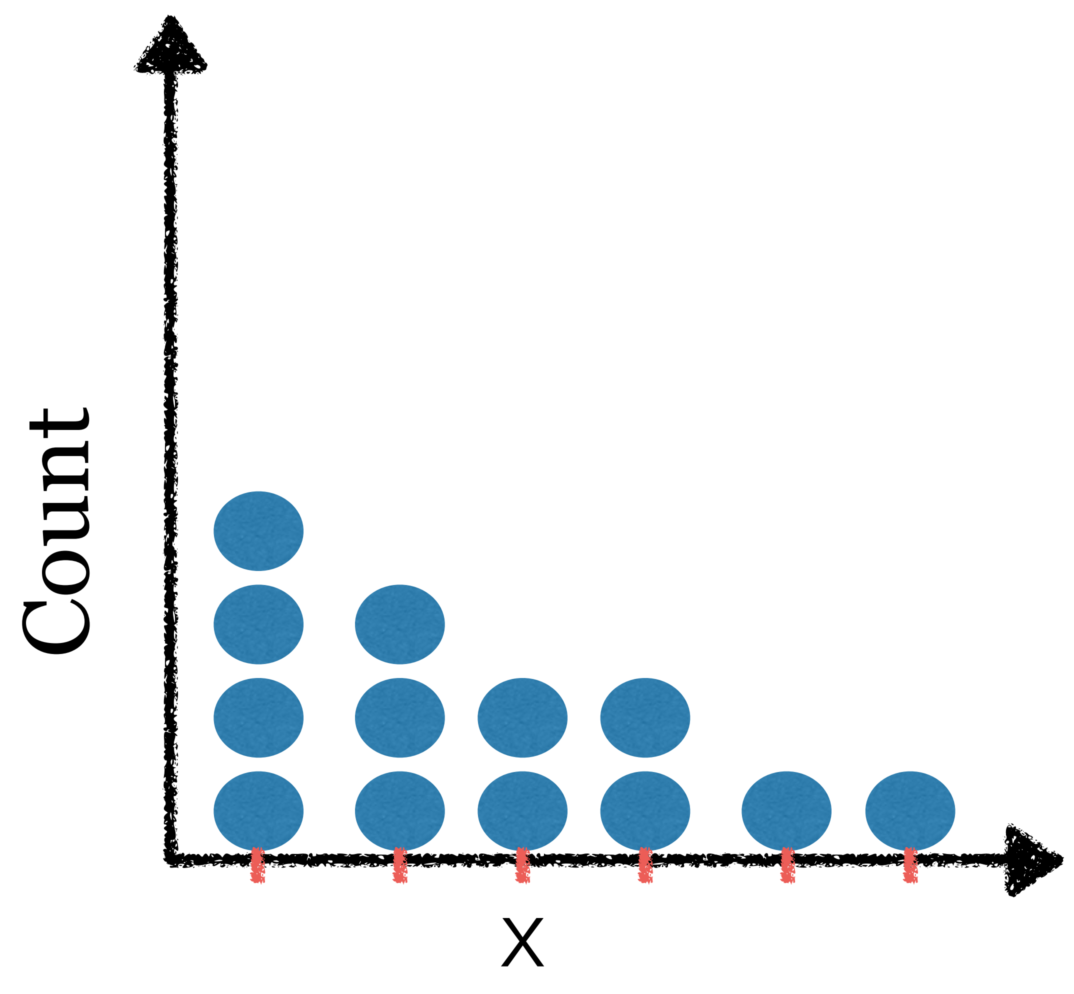
1-D histogram with 6 bins: few events enough to populate it



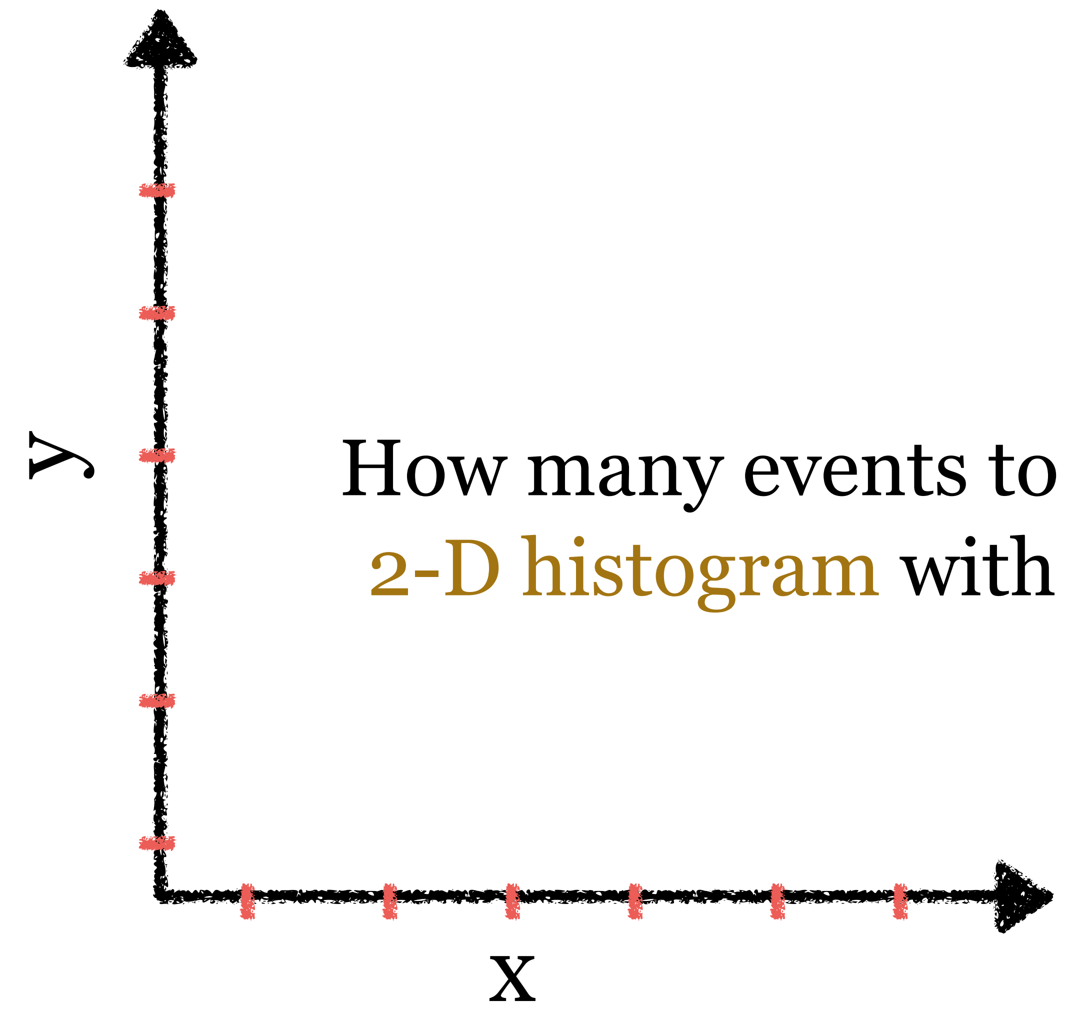
How many events to populate 2-D histogram with 6^2 bins ?

How many events for 50-D histogram with 6^{50} bins ?

But probability density estimation in higher dimensions is hard...



1-D histogram with 6 bins: few events enough to populate it



How many events to populate 2-D histogram with 6^2 bins ?

How many events for 50-D histogram with 6^{50} bins ?

Is there a better way to scale to higher dimensions ?

$\mathscr{D}$ back to being high-dimensional unbinned data
x_i an individual event

Foundational NSBI work: Cranmer et al: arXiv:1506.02169,
https://simulation-based-inference.org/

Neural networks can give us unbinned likelihood ratios

$$\mathscr{L}(\mu \,|\, \mathscr{D}) = p(\mathscr{D} \,|\, \mu)$$

Neyman–Pearson lemma: Likelihood ratio is the most powerful test statistic

We want to compare likelihoods:

$$\frac{p(\mathscr{D} \,|\, \mu)}{p(\mathscr{D} \,|\, ref)}$$

A neural network classifier trained on **simulated samples from** μ_1 **vs simulated samples from** *ref*, estimates the decision function:

$$s(x_i) = \frac{p(x_i \,|\, \mu_1)}{p(x_i \,|\, \mu_1) + p(x_i \,|\, ref)}$$

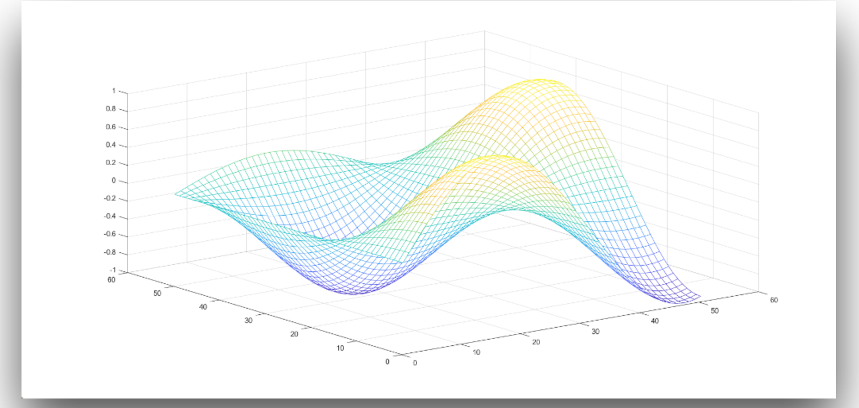
Which contains all the information required for the likelihood ratio:

$$\frac{p(x_i \,|\, \mu_1)}{p(x_i \,|\, ref)} = \frac{s(x_i)}{1 - s(x_i)}$$

- Optimal statistic to test each value of μ
- We get the LR *per event* (unbinned)

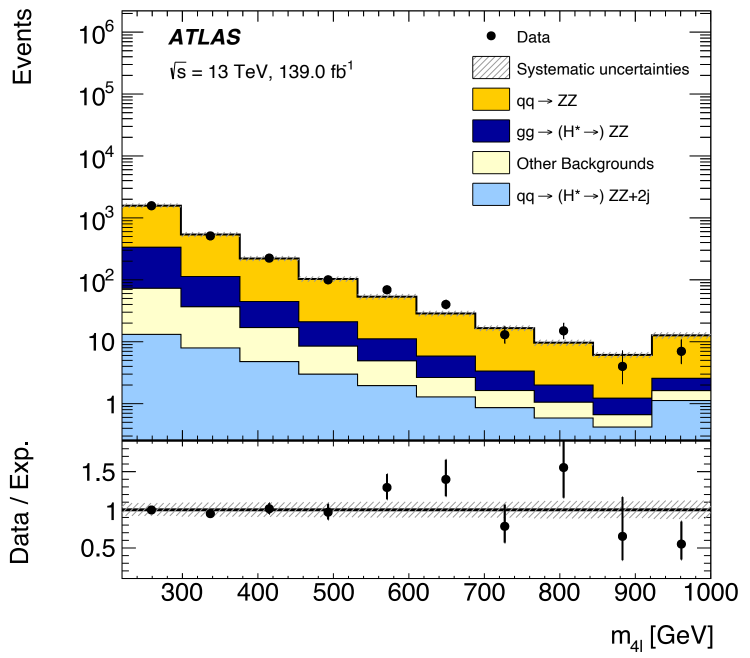
A new paradigm: Neural simulation-based inference (NSBI)

Traditional framework:



High-dim data

Summarisation to histogram

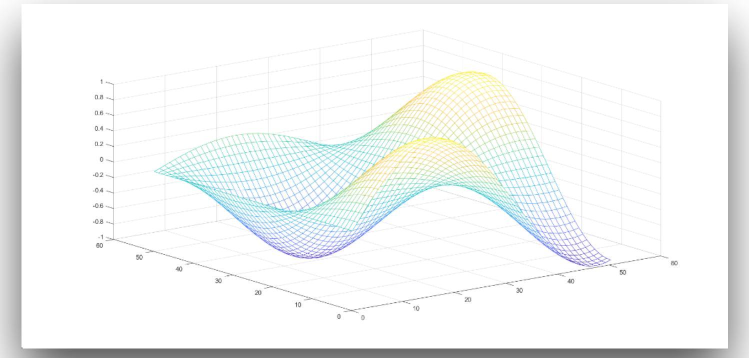


Summary Histogram → Statistical Fit → Likelihood $\mathscr{L}(\mu_1 | \mathscr{D})$

μ_1 → Statistical Fit

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



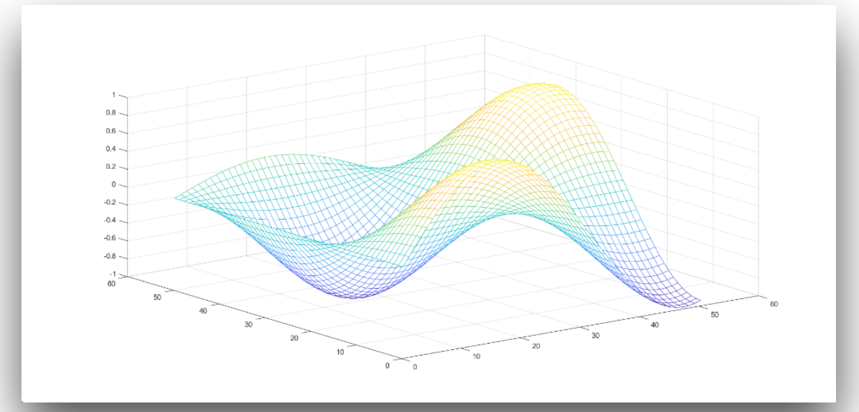
High-dim data

Obs Data → Neural Network → Likelihood Ratio $\left(\frac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})}\right)$

μ_1 → Neural Network

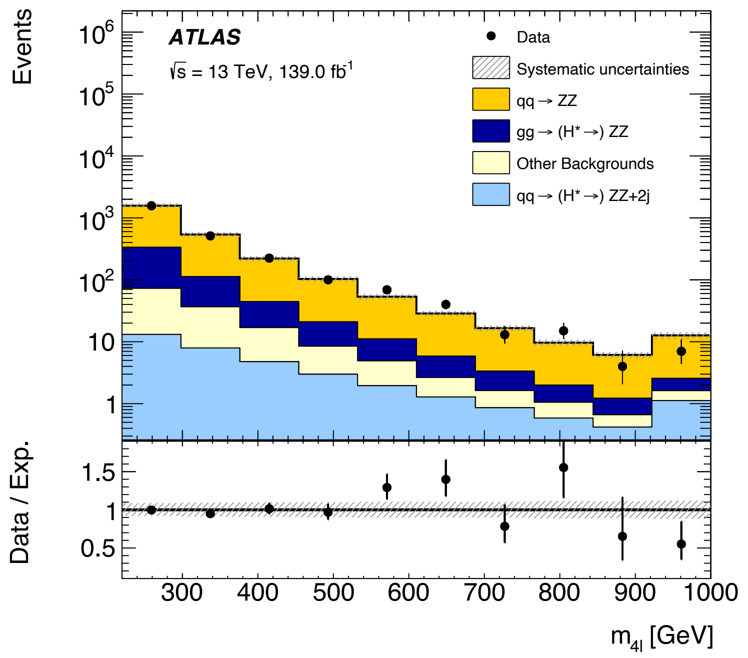
A new paradigm: Neural simulation-based inference (NSBI)

Traditional framework:



High-dim data

Summarisation to histogram



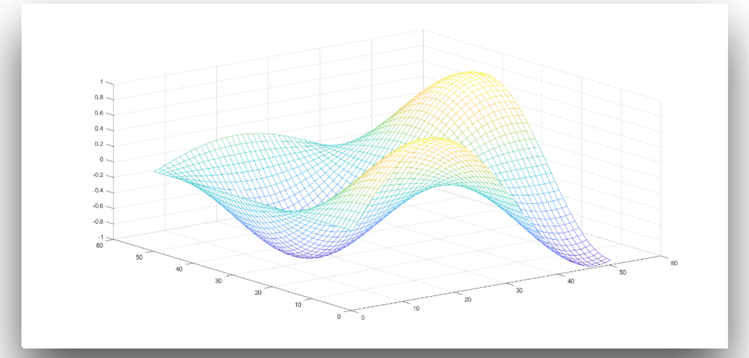
Summary Histogram → Statistical Fit → Likelihood $\mathscr{L}(\mu_1 | \mathscr{D})$

μ_1 → Statistical Fit

Hypothesis μ_1

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



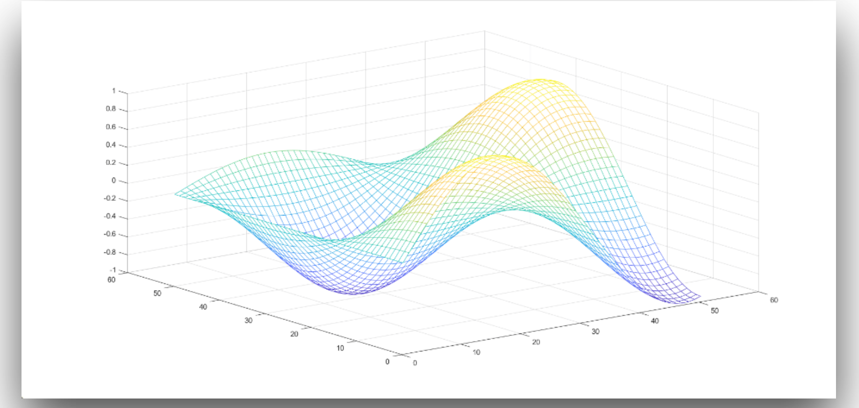
High-dim data

Obs Data → Neural Network → Likelihood Ratio $\left(\frac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})}\right)$

μ_1 → Neural Network

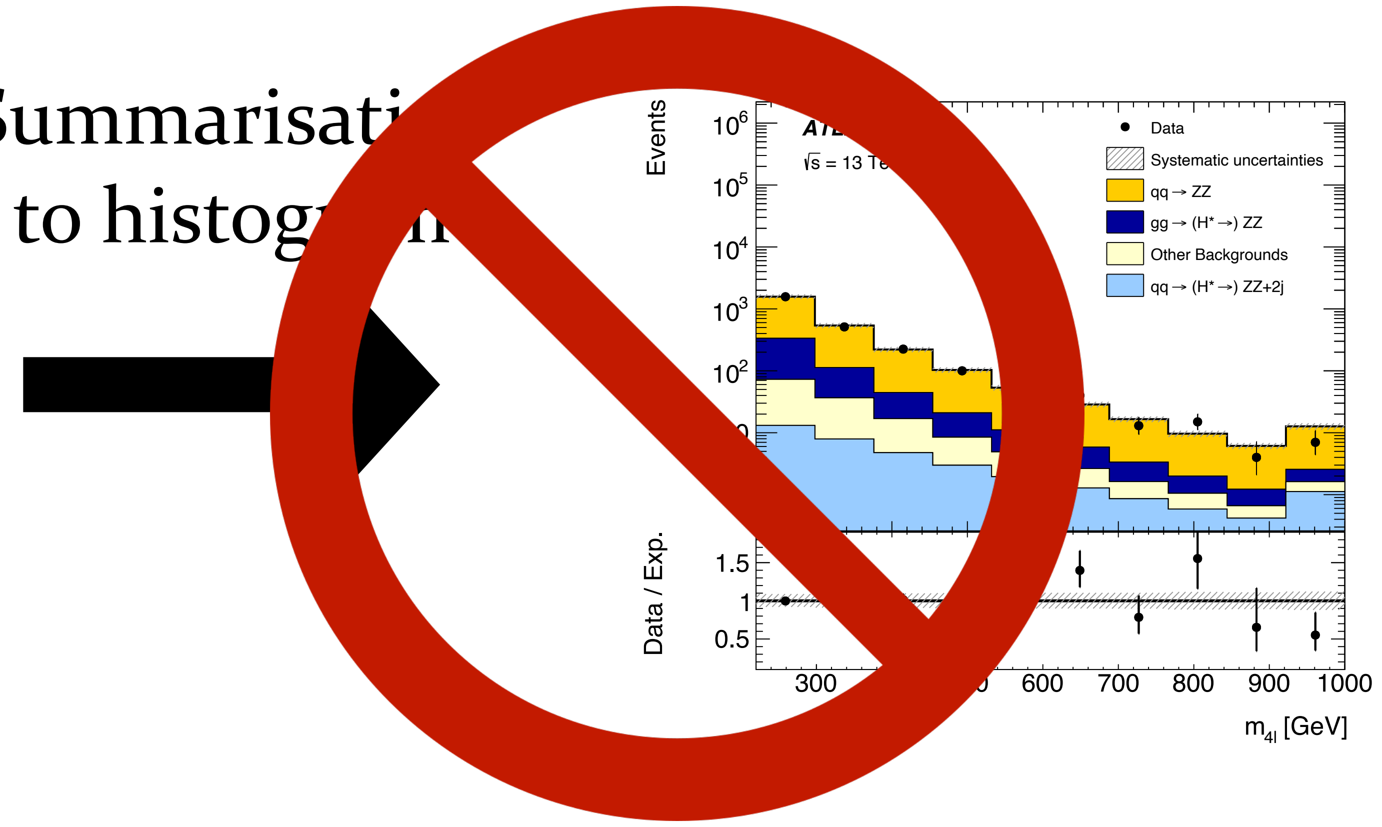
A new paradigm: Neural simulation-based inference (NSBI)

Traditional framework:



High-dim data

Summarisation to histogram



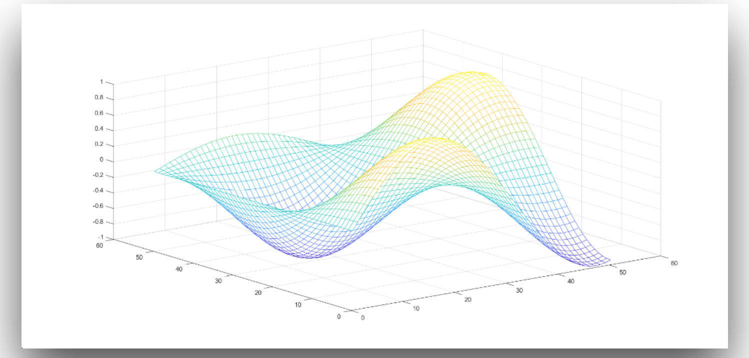
Summary Histogram → Statistical Fit → Likelihood $\mathscr{L}(\mu_1 | \mathscr{D})$

μ_1 → Statistical Fit

Hypothesis μ_1

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



High-dim data

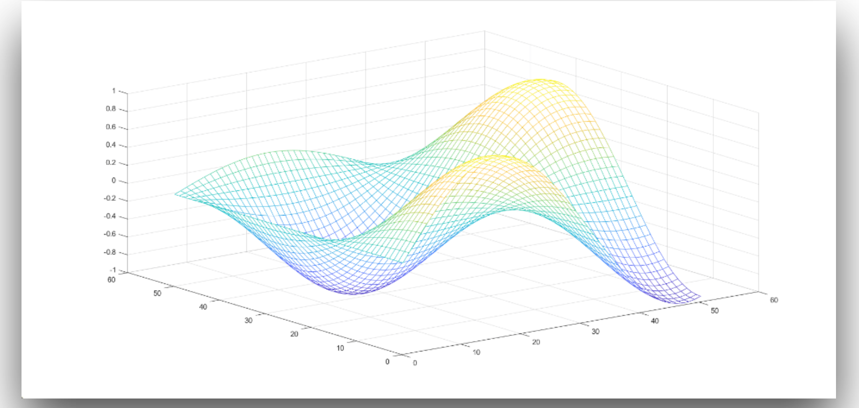
Obs Data → Neural Network → Likelihood Ratio $\left(\frac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})}\right)$

μ_1 → Neural Network

Foundational NSBI work: Cranmer et al: arXiv:1506.02169, https://simulation-based-inference.org/

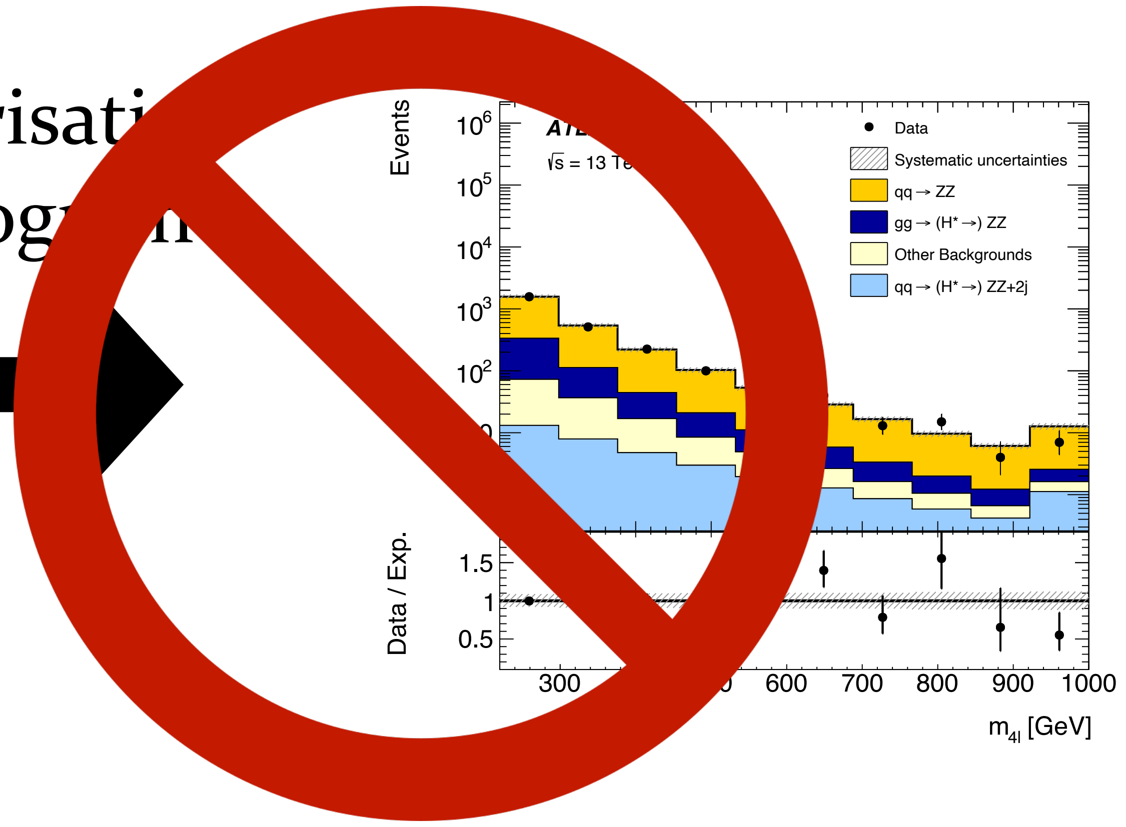
A new paradigm: Neural simulation-based inference (NSBI)

Traditional framework:



High-dim data

Summarisation to histogram



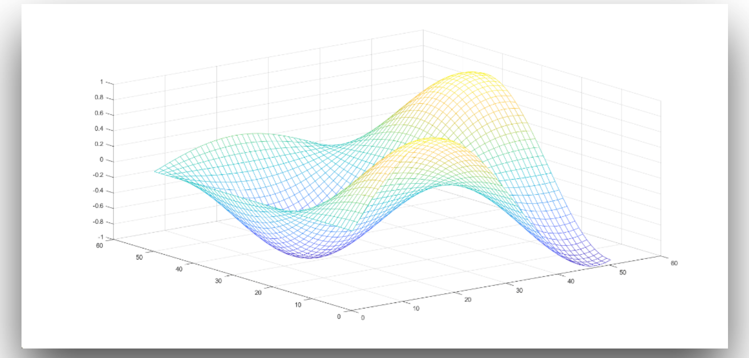
Summary Histogram → Statistical Fit → Likelihood $\mathscr{L}(\mu_1 | \mathscr{D})$

μ_1 → Statistical Fit

Hypothesis μ_1

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



High-dim data

Obs Data → Neural Network → Likelihood Ratio $\left(\frac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})}\right)$

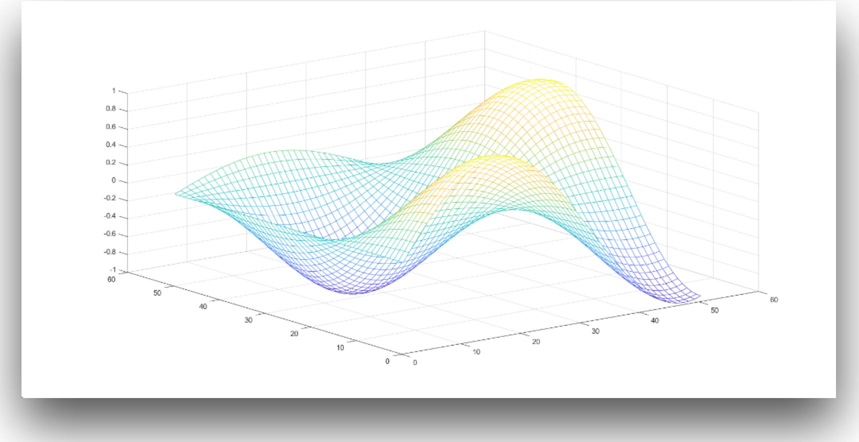
μ_1 → Neural Network

Hypothesis μ_1

Foundational NSBI work: Cranmer et al: arXiv:1506.02169, https://simulation-based-inference.org/

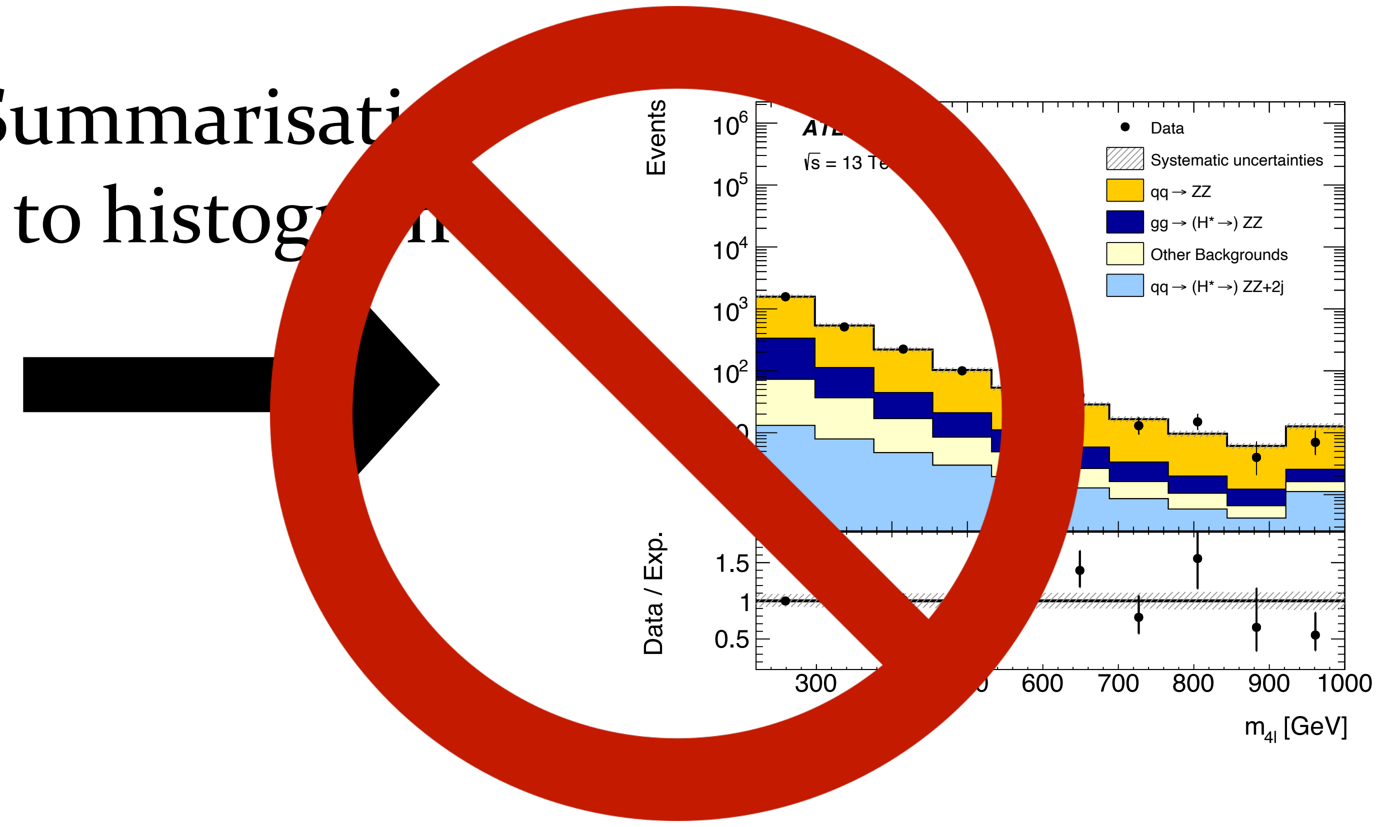
A new paradigm: Neural simulation-based inference (NSBI)

Traditional framework:



High-dim data

Summarisati
to histog



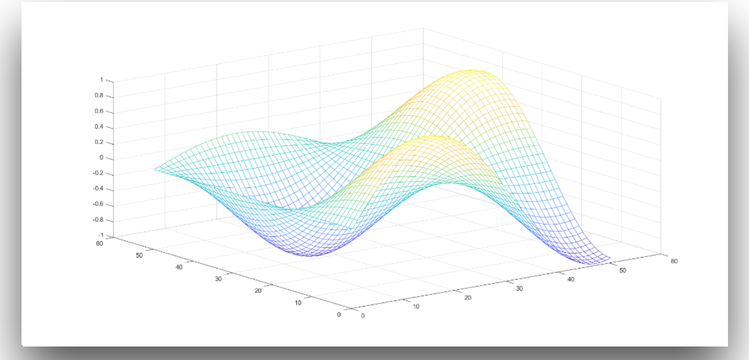
Summary Histogram → Statistical Fit → Likelihood $\mathscr{L}(\mu_1 | \mathscr{D})$

μ_1 → Statistical Fit

Hypothesis μ_1

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



High-dim data

Obs Data → Neural Network → Likelihood Ratio $\left(\frac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})}\right)$

μ_1 → Neural Network

Hypothesis μ_1

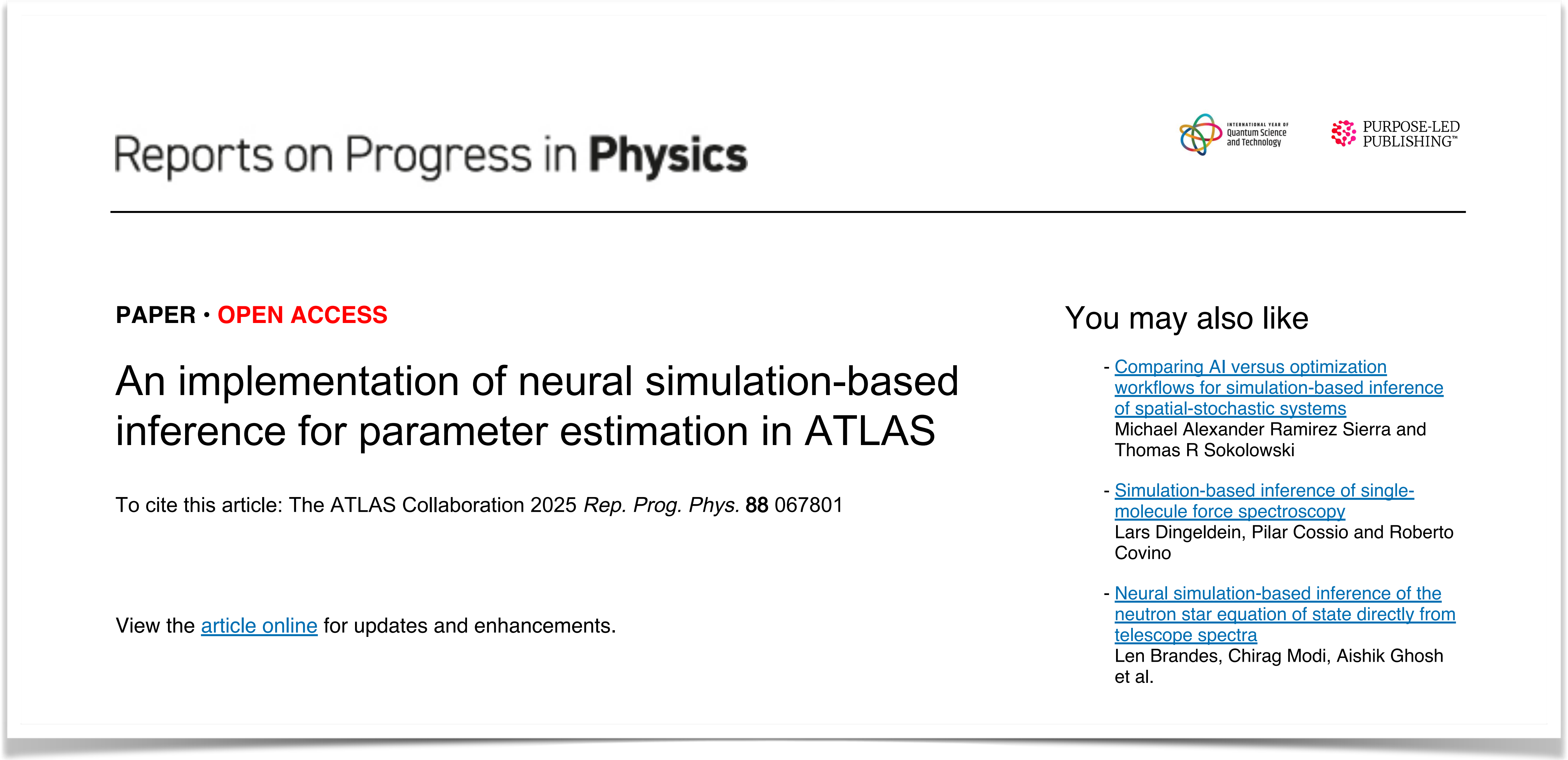
Train on simulations, apply on data

Is this dream attainable in practice?

- How to ensure robustness? By design and validation
- How to incorporate systematic uncertainties?
- How to perform Neyman construction in high dimensions?

Is this dream attainable in practice?

Yes!



Reports on Progress in **Physics**

PURPOSE-LED PUBLISHING

PAPER · **OPEN ACCESS**

An implementation of neural simulation-based inference for parameter estimation in ATLAS

To cite this article: The ATLAS Collaboration 2025 *Rep. Prog. Phys.* **88** 067801

View the article online for updates and enhancements.

You may also like

- Comparing AI versus optimization workflows for simulation-based inference of spatial-stochastic systems
  Michael Alexander Ramirez Sierra and Thomas R Sokolowski
- Simulation-based inference of single-molecule force spectroscopy
  Lars Dingeldein, Pilar Cossio and Roberto Covino
- Neural simulation-based inference of the neutron star equation of state directly from telescope spectra
  Len Brandes, Chirag Modi, Aishik Ghosh et al.

Is this dream attainable in practice?

Yes!

Reports on Progress in **Physics**

PAPER · OPEN ACCESS

An implementation of neural sin
inference for parameter estimat

To cite this article: The ATLAS Collaboration 2025 *Rep. Prog.*

View the article online for updates and enhancements.

Reports on Progress in **Physics**

PAPER · OPEN ACCESS

Measurement of off-shell Higgs boson production in the $H^* \to ZZ \to 4\ell$ decay channel using a neural simulation-based inference technique in 13 TeV *pp* collisions with the ATLAS detector

To cite this article: The ATLAS Collaboration 2025 *Rep. Prog. Phys.* **88** 057803

View the article online for updates and enhancements.

You may also like

- Dark matter search with a resonantly-coupled hybrid spin system
Kai Wei, Zitong Xu, Yuxuan He et al.
- Beyond Kitaev physics in strong spin-orbit coupled magnets
Ioannis Rousochatzakis, Natalia B Perkins, Qiang Luo et al.
- 70 years of hyperon spectroscopy: a review of strange , baryons, and the spectrum of charmed and bottom baryons
Volker Crede and John Yelton

Applied on Run2 data, superseding previous ATLAS paper on same data !

ars TECHNICA

AI BIZ & IT CARS CULTURE GAMING HEALTH POLICY SCIENCE SECURITY SPACE TECH FORUM SUBSCRIBE SIGN IN

A VERY FUN PROBLEM

How a grad student got LHC data to play nice with quantum interference

New approach is already having an impact on the experiment's plans for future work.

MATT VON HIPPEL – 23 JUN 2025 11:00 70

Is this dream attainable in practice?

Yes!

Reports on Progress in **Physics**

PAPER • OPEN ACCESS

An implementation of neural sin
inference for parameter estimat

To cite this article: The ATLAS Collaboration 2025 *Rep. Prog.*

View the article online for updates and enhancements.

Reports on Progress in **Physics**

PAPER • OPEN ACCESS

Measurement of off-shell Higgs boson production in the $H^* \to ZZ \to 4\ell$ decay channel using a neural simulation-based inference technique in 13 TeV *pp* collisions with the ATLAS detector

To cite this article: The ATLAS Collaboration 2025 *Rep. Prog. Phys.* **88** 057803

View the article online for updates and enhancements.

You may also like

- Dark matter search with a resonantly-coupled hybrid spin system
Kai Wei, Zitong Xu, Yuxuan He et al.
- Beyond Kitaev physics in strong spin-orbit coupled magnets
Ioannis Rousochatzakis, Natalia B Perkins, Qiang Luo et al.
- 70 years of hyperon spectroscopy: a review of strange , baryons, and the spectrum of charmed and bottom baryons
Volker Crede and John Yelton

Applied on Run2 data, superseding previous ATLAS paper on same data !

ars TECHNICA AI BIZ & IT CARS CULTURE GAMING HEALTH POLICY SCIENCE SECURITY SPACE TECH FORUM SUBSCRIBE SIGN IN

A VERY FUN PROBLEM

How a grad student got LHC data to play nice with quantum interference

New approach is already having an impact on the experiment's plans for future work.

MATT VON HIPPEL – 23 JUN 2025 11:00 70

Today's tutorial will follow this approach, although many variations are possible

Is this dream attainable in practice?

Yes!

Reports on Progress in **Physics**

PURPOSE-LED PUBLISHING™

PAPER • OPEN ACCESS

An implementation of neural sin
inference for parameter estimat

To cite this article: The ATLAS Collaboration 2025 *Rep. Prog.*

View the article online for updates and enhancements.

Reports on Progress in **Physics**

PURPOSE-LED PUBLISHING™

PAPER • OPEN ACCESS

Measurement of off-shell Higgs boson production in the $H^* \to ZZ \to 4\ell$ decay channel using a neural simulation-based inference technique in 13 TeV *pp* collisions with the ATLAS detector

To cite this article: The ATLAS Collaboration 2025 *Rep. Prog. Phys.* **88** 057803

View the article online for updates and enhancements.

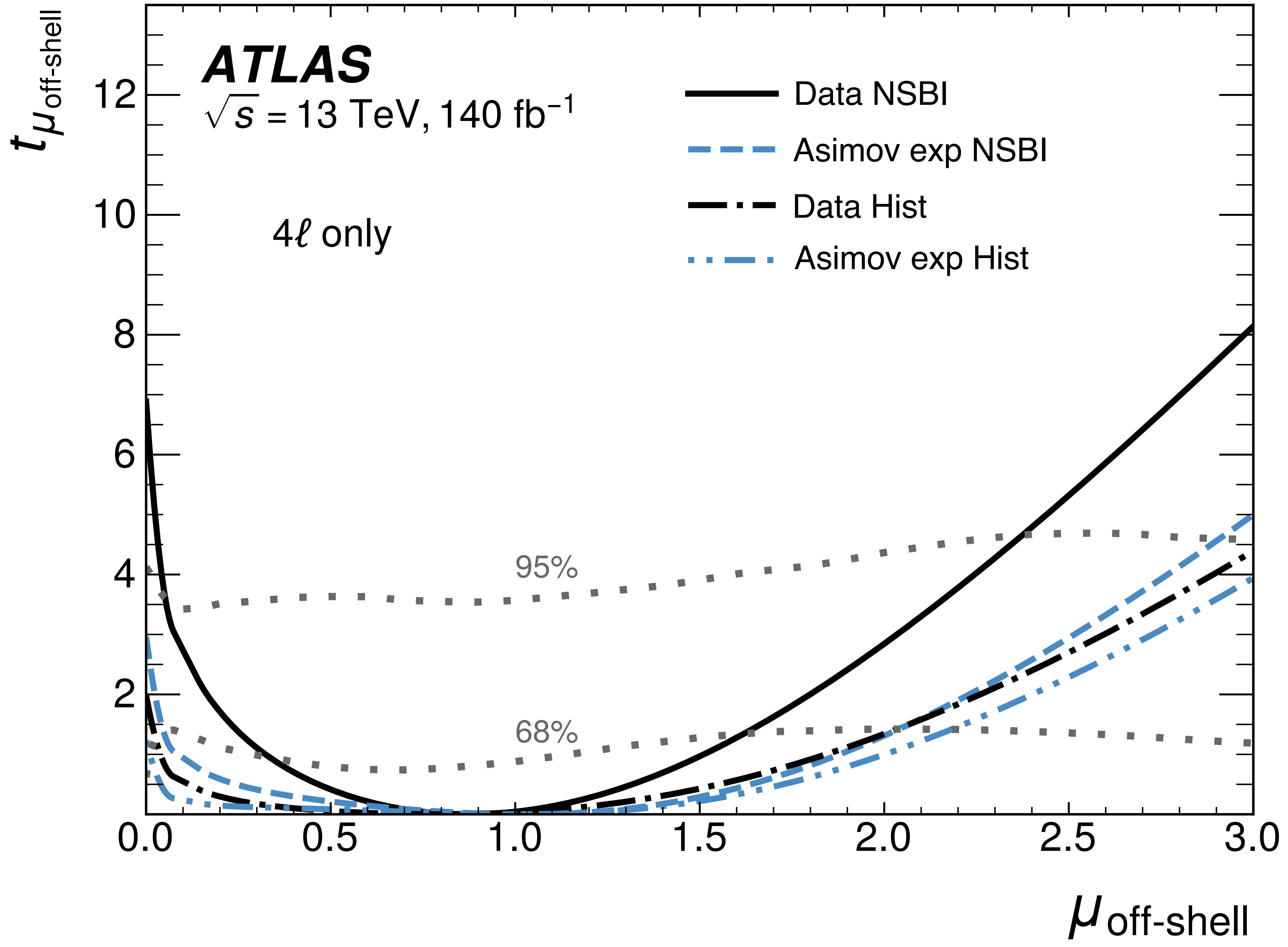
You may also like

- Dark matter search with a resonantly-coupled hybrid spin system
Kai Wei, Zitong Xu, Yuxuan He et al.
- Beyond Kitaev physics in strong spin-orbit coupled magnets
Ioannis Rousochatzakis, Natalia B Perkins, Qiang Luo et al.
- 70 years of hyperon spectroscopy: a review of strange , baryons, and the spectrum of charmed and bottom baryons
Volker Crede and John Yelton

Applied on Run2 data, superseding previous ATLAS paper on same data !

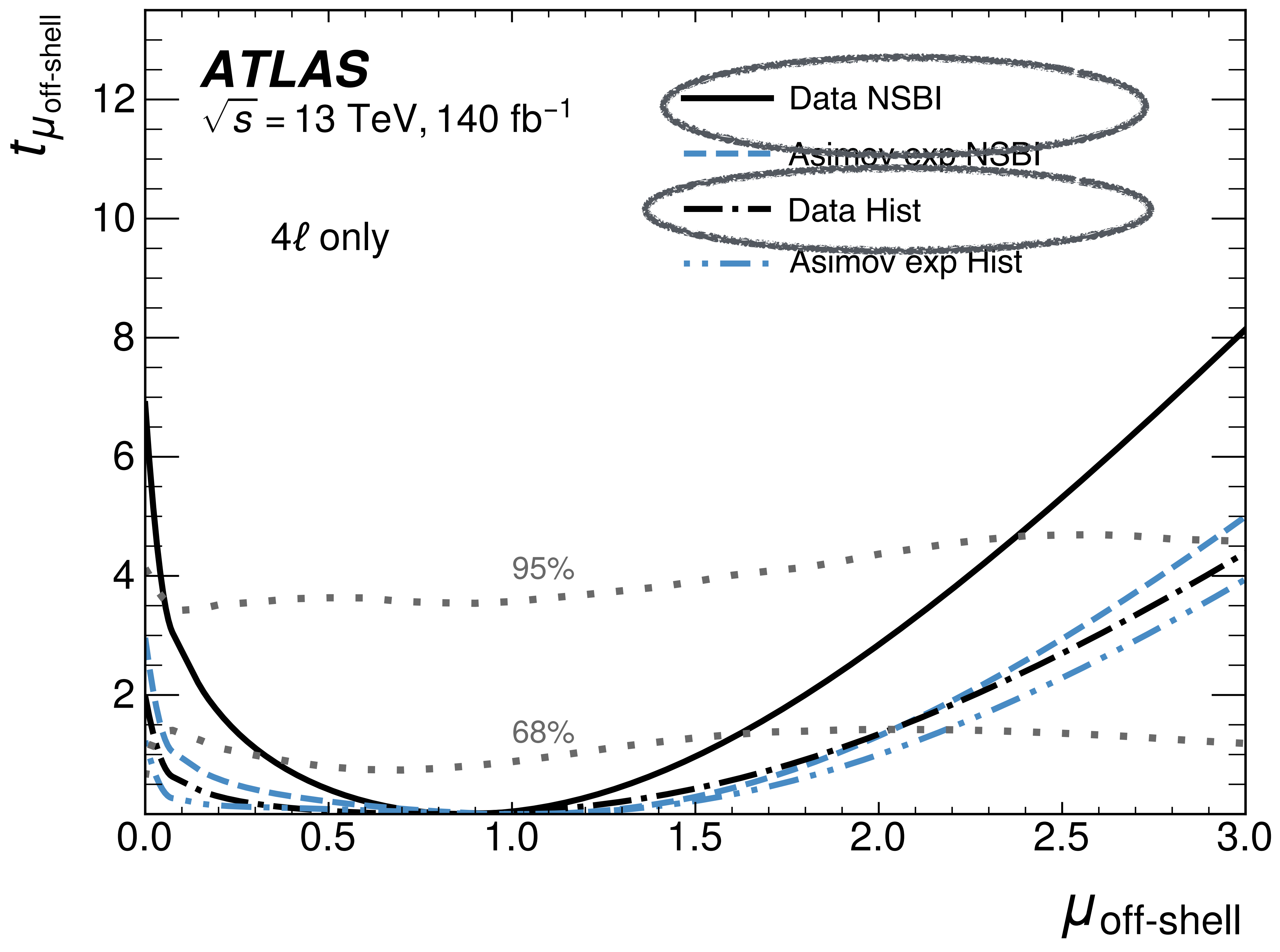
Application in off-shell Higgs Run 2 analysis: big improvement!

NSBI vs histogram analysis



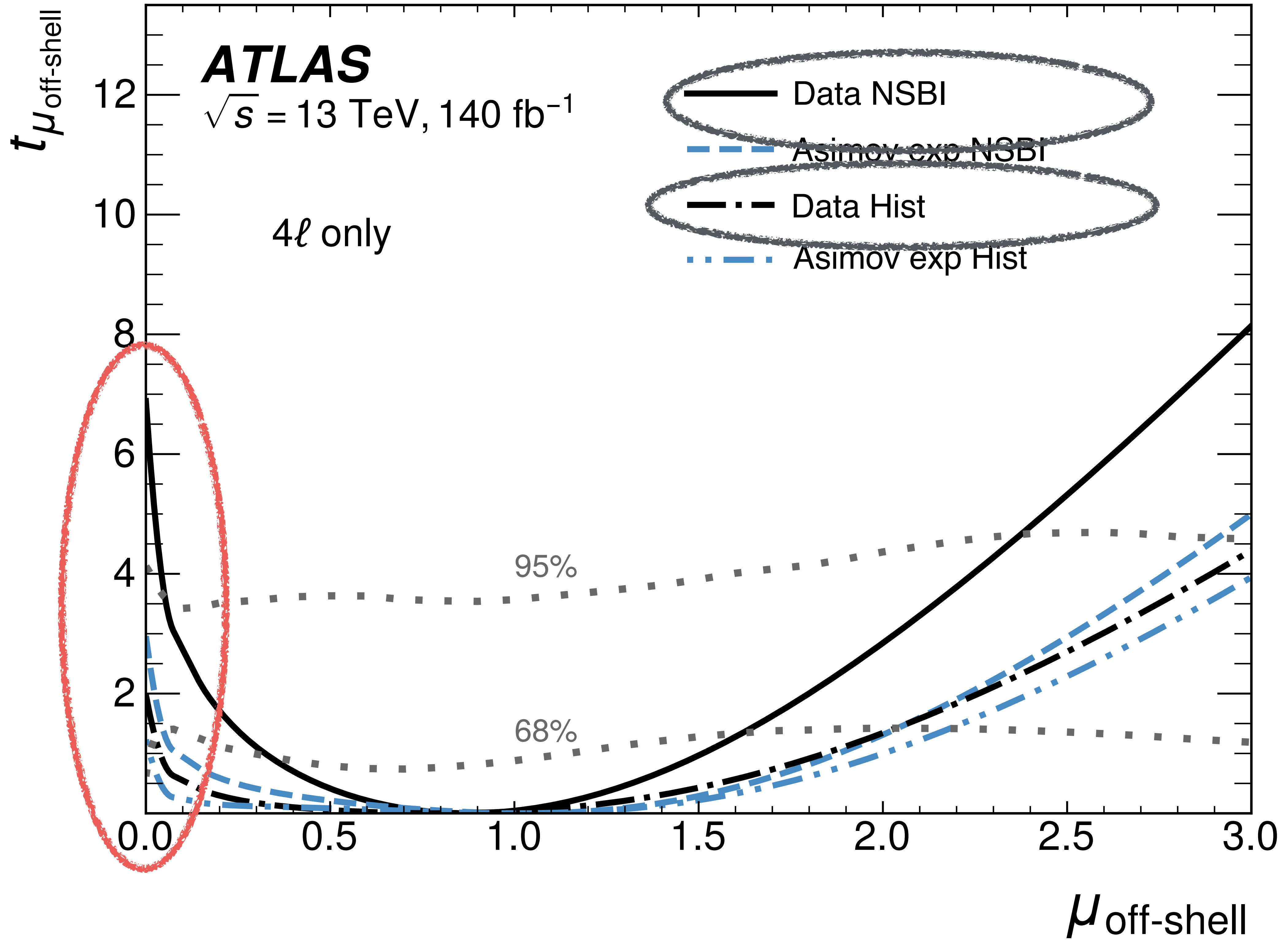
Application in off-shell Higgs Run 2 analysis: big improvement!

NSBI vs histogram analysis



Application in off-shell Higgs Run 2 analysis: big improvement!

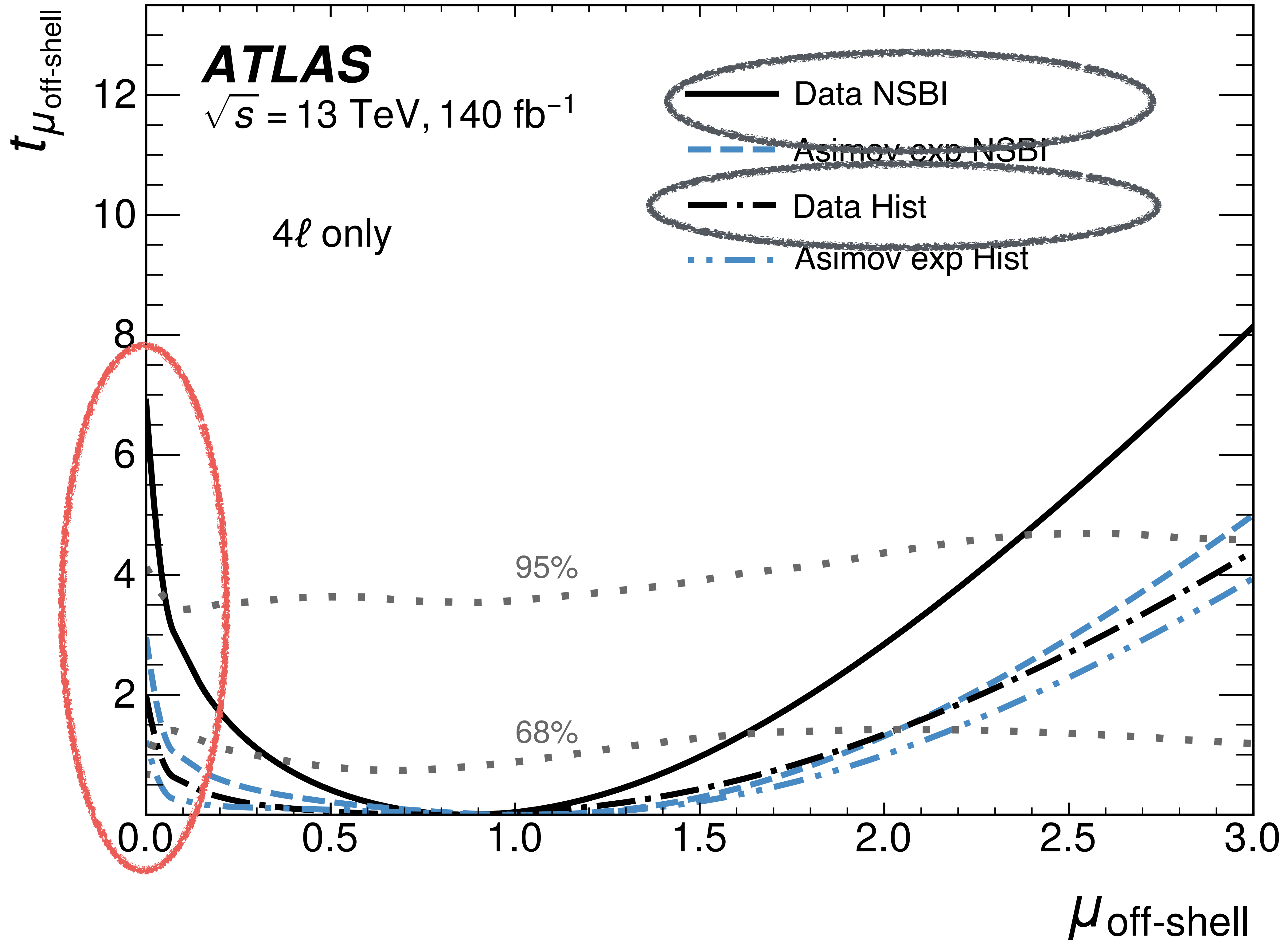
NSBI vs histogram analysis



Unprecedented improvement in ability to reject null hypothesis!

Application in off-shell Higgs Run 2 analysis: big improvement!

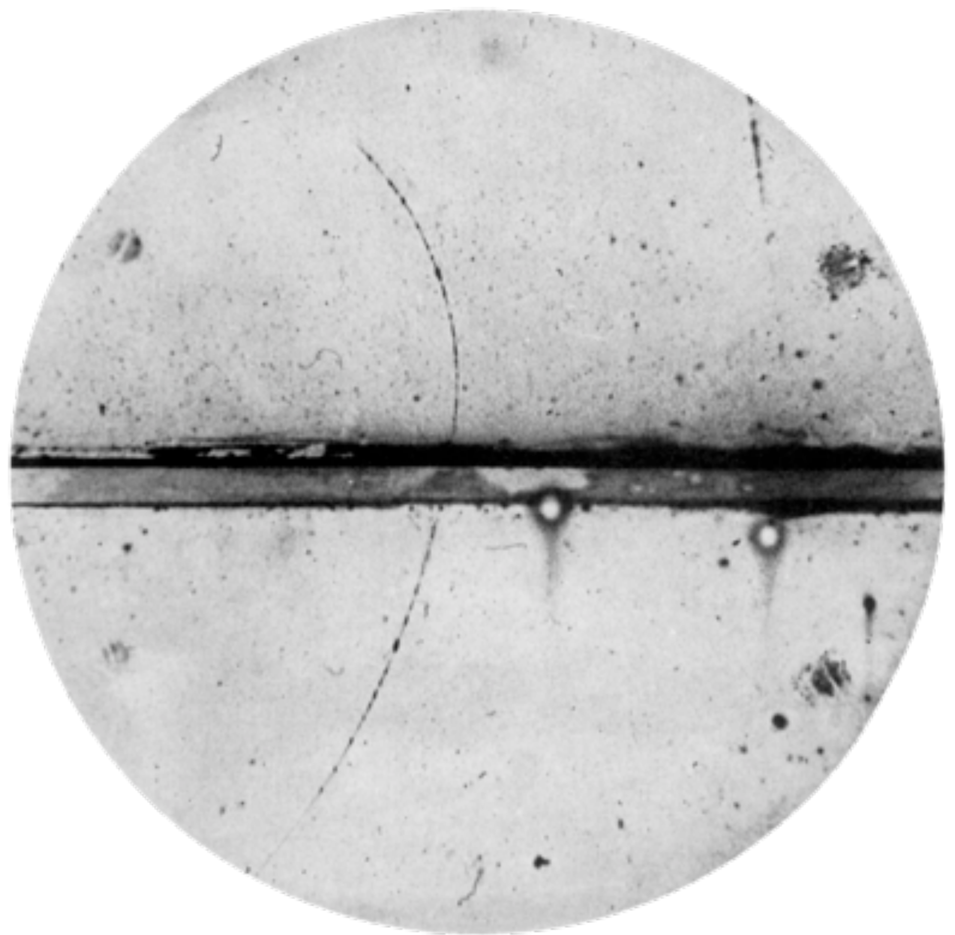
NSBI vs histogram analysis



Unprecedented improvement in ability to reject null hypothesis!

Observed data happens to provide stronger constraint than Asimov for both hist and NSBI

Positron discovery (1930s)



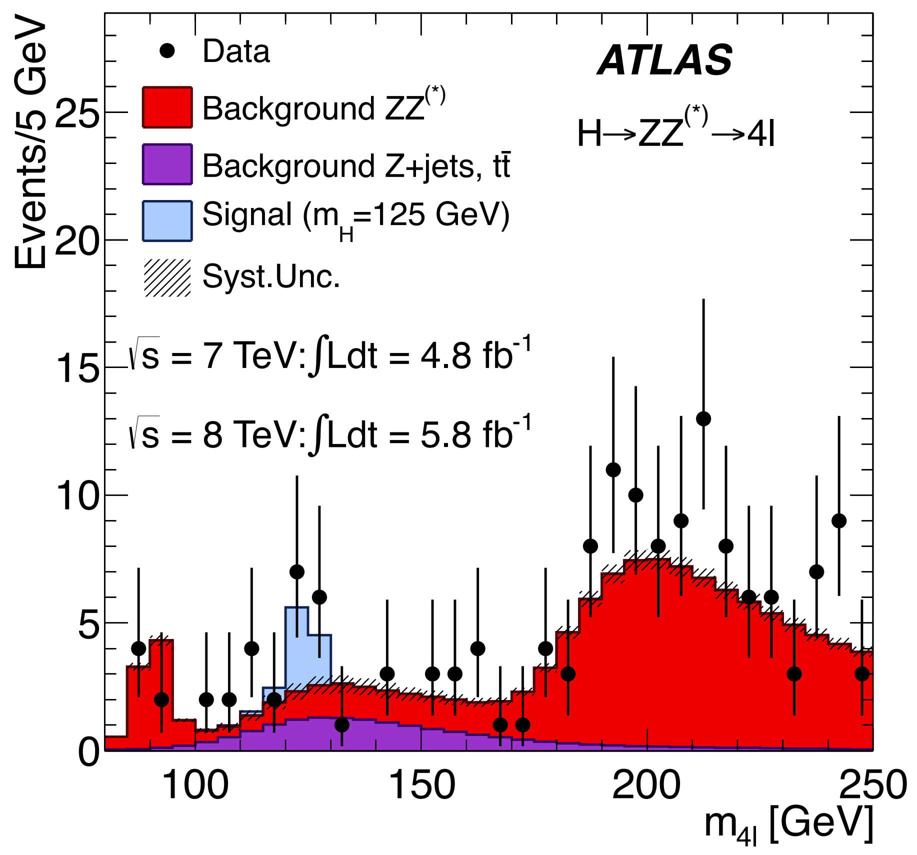
Single event

Top quark discovery (1990s)

Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}

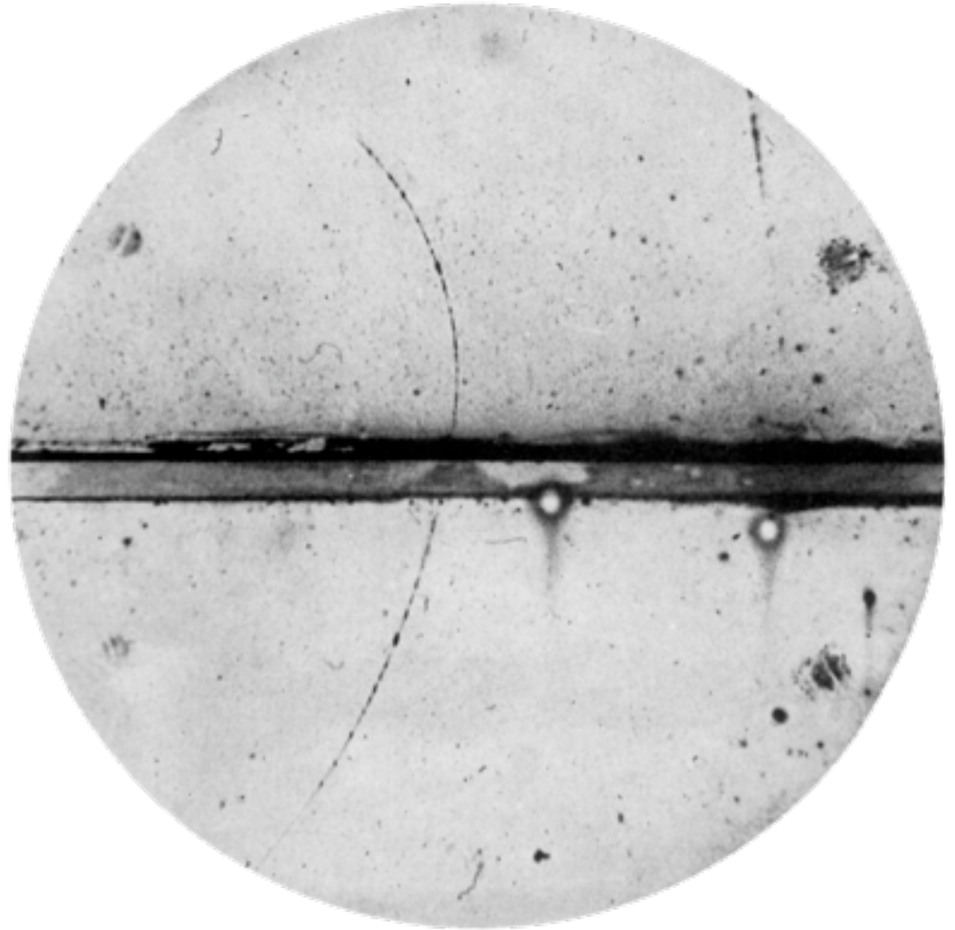
Multiple events:
Cut-and-count

Higgs boson
discovery (2010s)



Shape information:
Histogram

Positron discovery (1930s)



Single event

Image: Wikipedia / PhysRev.43.491

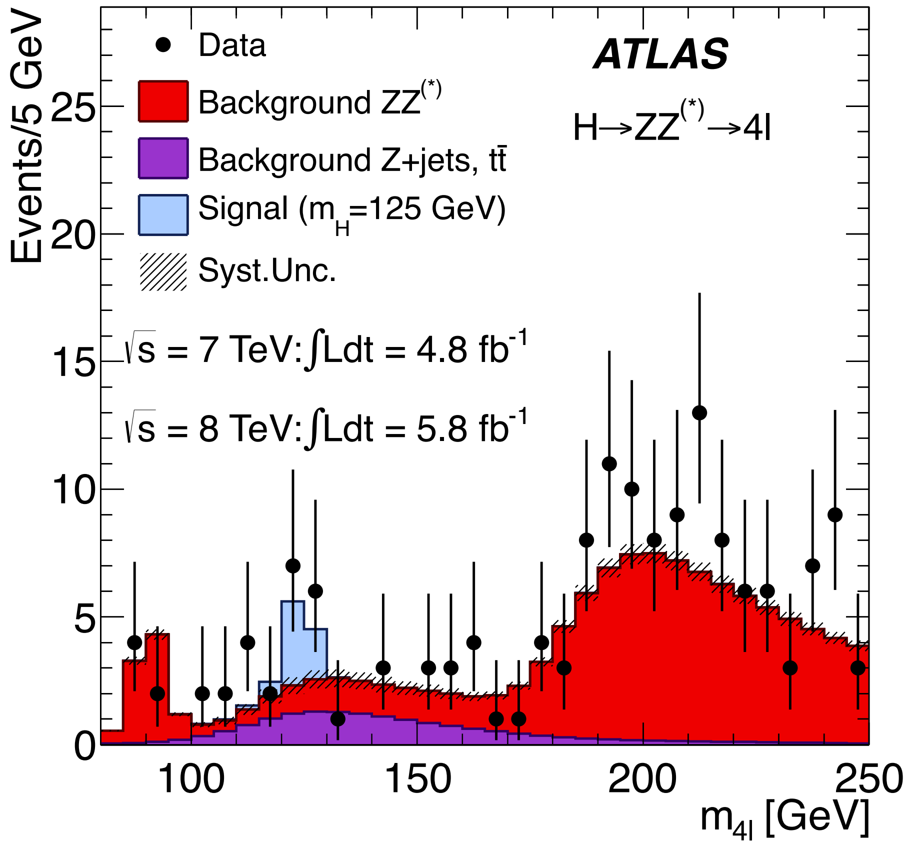
Top quark discovery (1990s)

Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}

Multiple events:
Cut-and-count

CDF Collaboration: arXiv:9503002

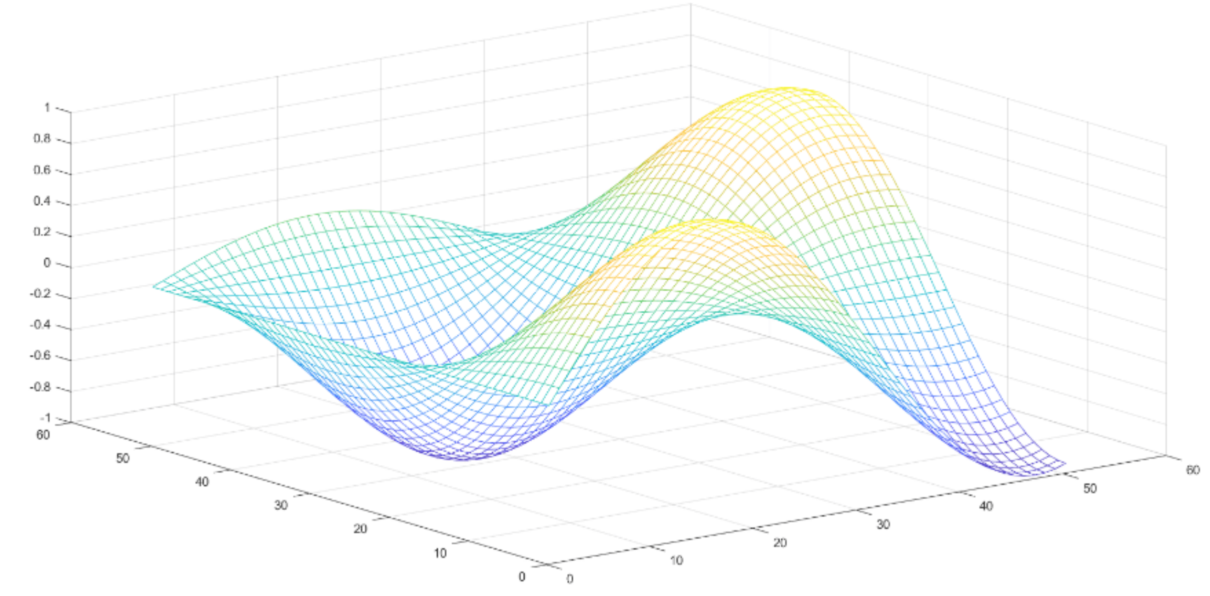
Higgs boson discovery (2010s)



Shape information:
Histogram

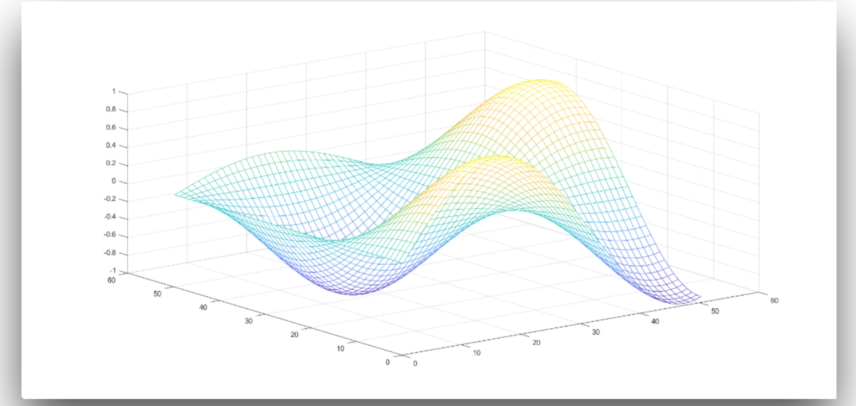
ATLAS Collaboration: arXiv:1207.7214

Future discovery (2020s ?)

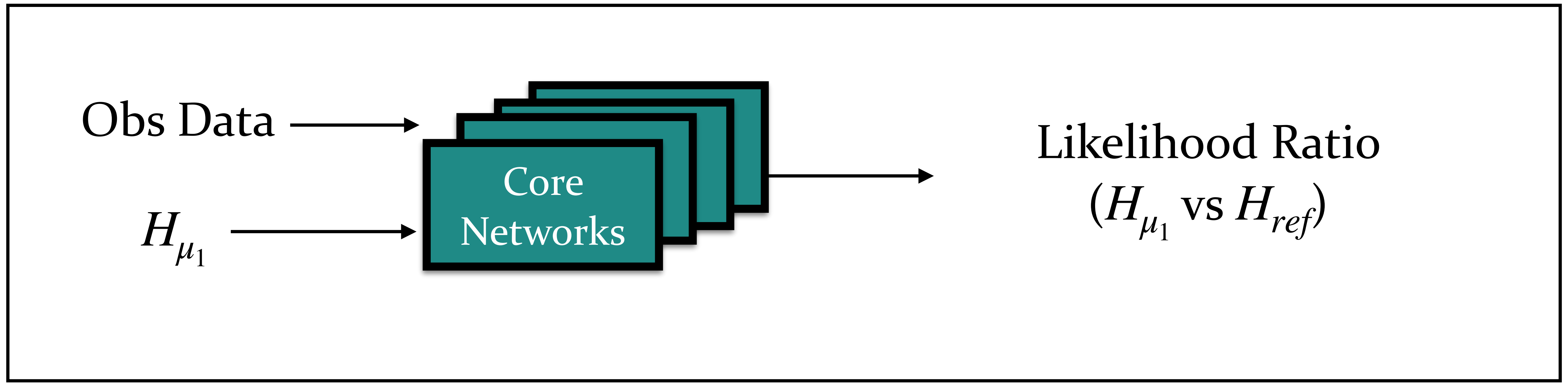


High-dim shape information, continuous (i.e. unbinned):
Neural inference

A robust NSBI implementation



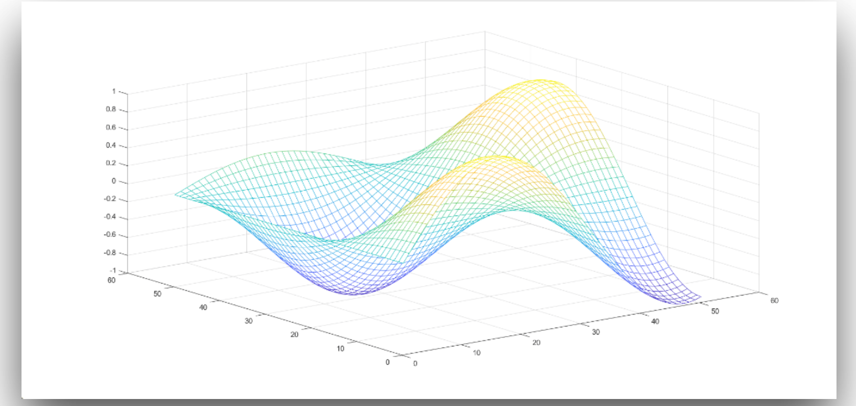
$O(16)$ observables



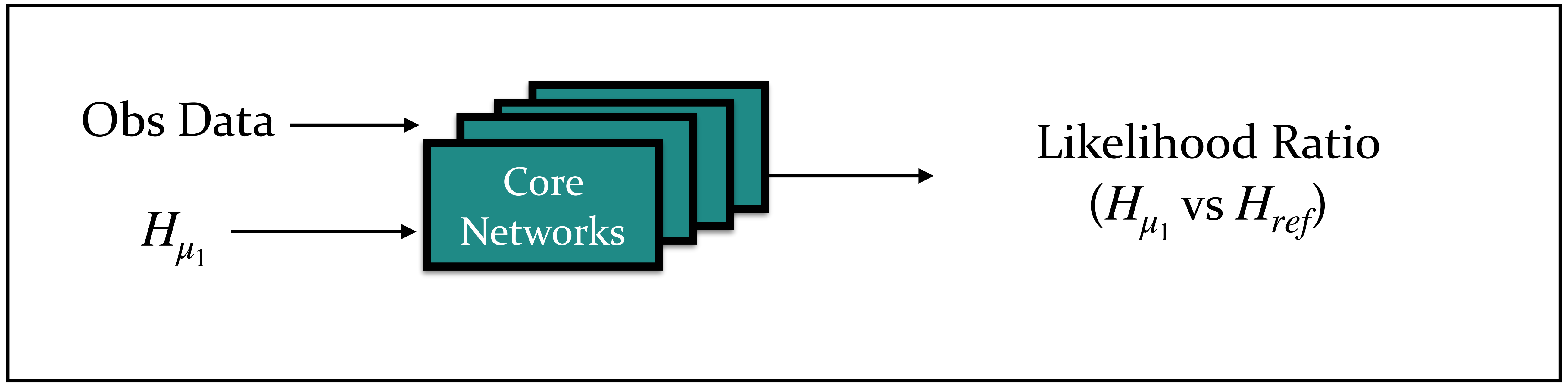
$$\frac{p(x_i \,|\, \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu, \alpha)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_{k}^{N_{syst}} G_j(\alpha_k) \cdot \frac{p_j(x_i, \alpha_k)}{p_j(x_i)}$$

ATLAS methods paper: arXiv:2412.01600

A robust NSBI implementation

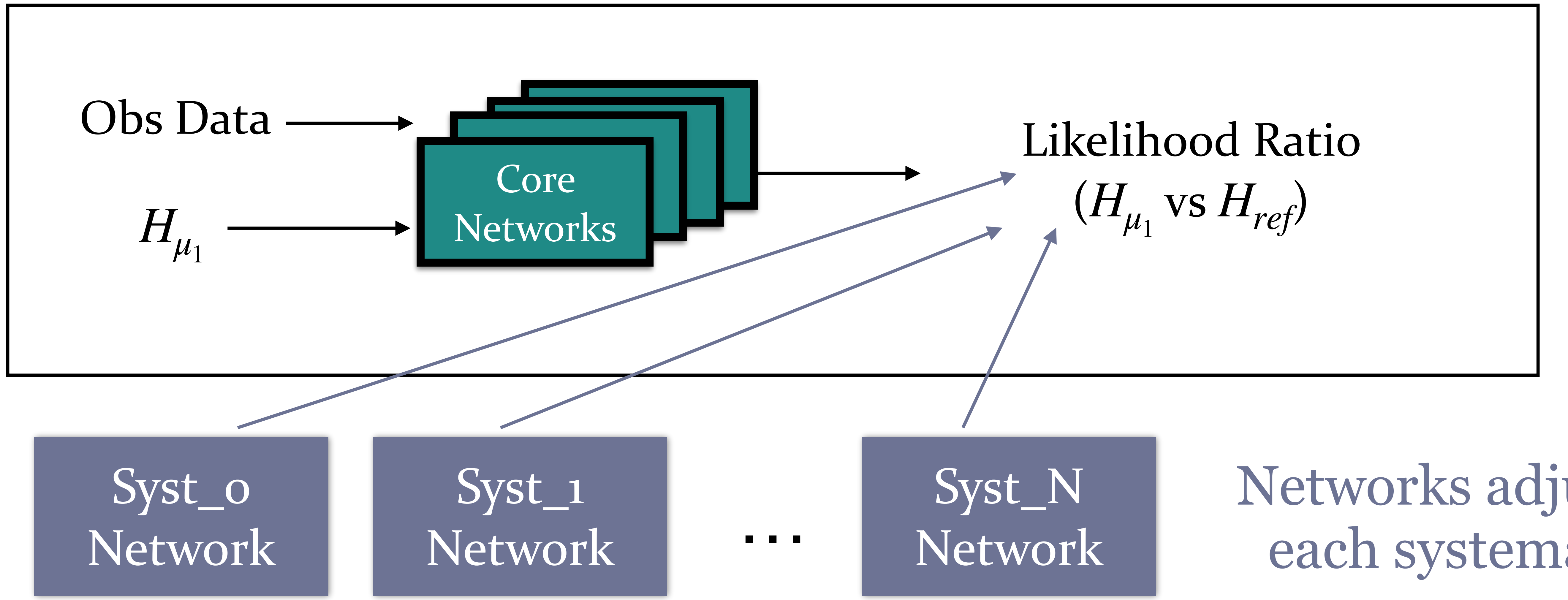


$O(16)$ observables



$$\frac{p(x_i\,|\,\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu,\alpha)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_{k}^{N_{syst}} G_j(\alpha_k) \cdot \frac{p_j(x_i,\alpha_k)}{p_j(x_i)}$$

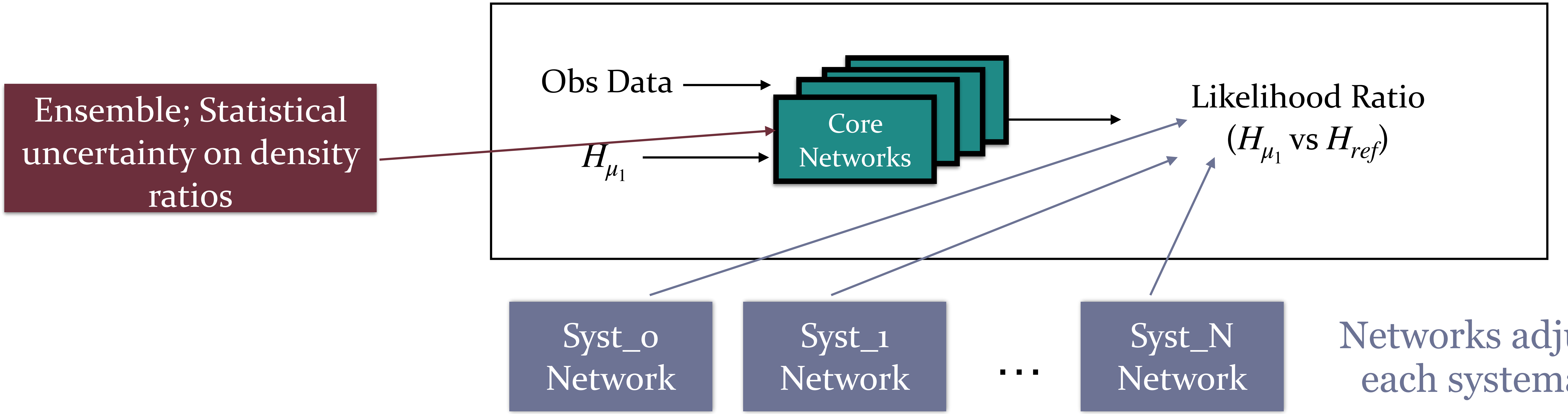
A robust NSBI implementation



Networks adjust likelihood for each systematic uncertainty

$$\frac{p(x_i\,|\,\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu,\alpha)} \sum_j^C f_j(\mu)\cdot \nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_k^{N_{syst}} G_j(\alpha_k)\cdot \frac{p_j(x_i,\alpha_k)}{p_j(x_i)}$$

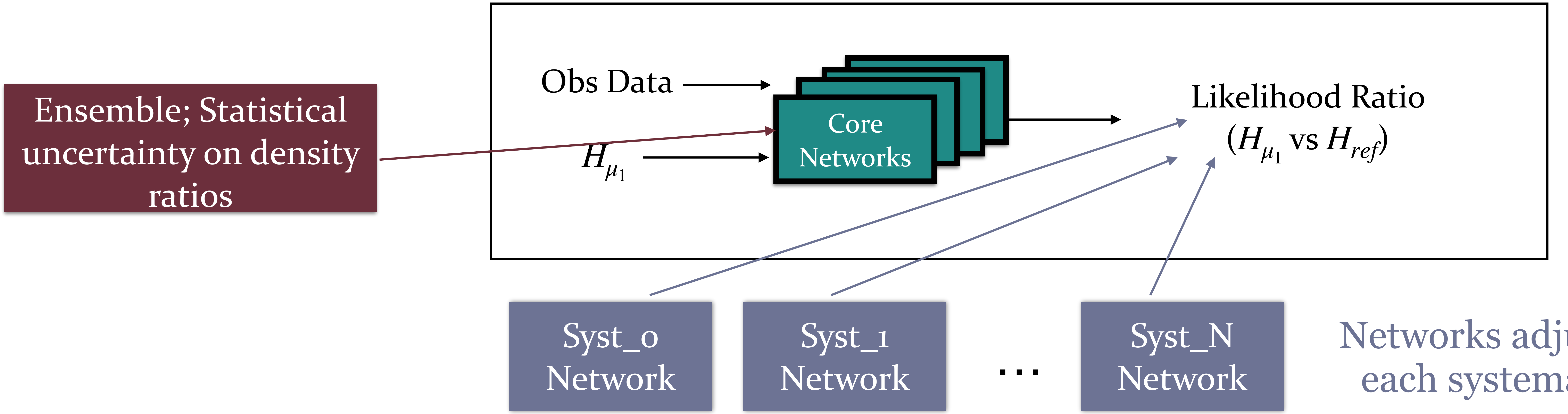
A robust NSBI implementation



Networks adjust likelihood for each systematic uncertainty

$$\frac{p(x_i\,|\,\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu,\alpha)} \sum_j^C f_j(\mu)\cdot\nu_j\cdot\frac{p_j(x_i)}{p_{ref}(x_i)}\cdot\prod_k^{N_{syst}} G_j(\alpha_k)\cdot\frac{p_j(x_i,\alpha_k)}{p_j(x_i)}$$

A robust NSBI implementation



Networks adjust likelihood for each systematic uncertainty

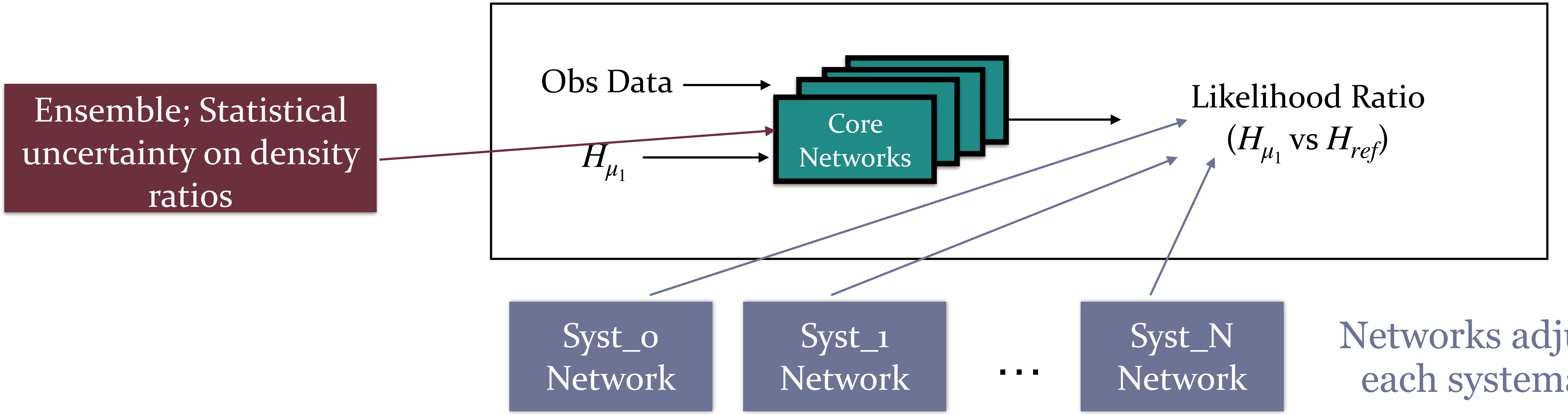
$$\frac{p(x_i\,|\,\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu,\alpha)}\sum_j^C f_j(\mu)\cdot\nu_j\cdot\frac{p_j(x_i)}{p_{ref}(x_i)}\cdot\prod_k^{N_{syst}} G_j(\alpha_k)\cdot\frac{p_j(x_i,\alpha_k)}{p_j(x_i)}$$

Training details

- ✦ Train $O(10^4)$ networks on TensorFlow
- ✦ Computing resources provided by Google, SMU, other HPC clusters
- ✦ Fits with JAX

A robust NSBI implementation

NN learns probability ratios in high dim



Networks adjust likelihood for each systematic uncertainty

$$\frac{p(x_i\,|\,\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu,\alpha)} \sum_{j}^{C} f_j(\mu)\cdot\nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_{k}^{N_{syst}} G_j(\alpha_k) \cdot \frac{p_j(x_i,\alpha_k)}{p_j(x_i)}$$

Training details

- ✦ Train $O(10^4)$ networks on TensorFlow
- ✦ Computing resources provided by Google, SMU, other HPC clusters
- ✦ Fits with JAX

Is this dream attainable in practice?

- How to ensure robustness? By design and validation
- How to incorporate systematic uncertainties?
- How to perform Neyman construction in high dimensions?

ATLAS methods paper: arXiv:2412.01600

Semi-parametric NSBI

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i)$$

j runs over different physics process
(Eg. $gg \rightarrow H^* \rightarrow 4l$, $gg \rightarrow ZZ \rightarrow 4l$)

Example use case

ATLAS methods paper: arXiv:2412.01600

Semi-parametric NSBI

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \; p_j(x_i)$$

j runs over different physics process (Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Example use case

$$p_{\text{ggF}}(x|\mu) = \frac{1}{\nu_{\text{ggF}}(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S \; p_S(x) + \sqrt{\mu}\, \nu_{\text{SBI}_1} \; p_{\text{SBI}_1}(x) + (1 - \sqrt{\mu}) \nu_{\text{B}} \; p_{\text{B}}(x) \right]$$

Semi-parametric NSBI

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i)$$

j runs over different physics process
(Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Comes from theory model chosen to interpret data

Example use case

$$p_{\text{ggF}}(x|\mu) = \frac{1}{\nu_{\text{ggF}}(\mu)} \left[\left(\mu - \sqrt{\mu}\right) \nu_S \, p_S(x) + \sqrt{\mu}\, \nu_{\text{SBI}_1} \, p_{\text{SBI}_1}(x) + \left(1 - \sqrt{\mu}\right) \nu_B \, p_B(x) \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

Semi-parametric NSBI

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i)$$

Event rates estimated from simulations

Comes from theory model chosen to interpret data

j runs over different physics process
(Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Example use case

$$p_{\text{ggF}}(x|\mu) = \frac{1}{\nu_{\text{ggF}}(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S \, p_S(x) + \sqrt{\mu}\, \nu_{\text{SBI}_1} \, p_{\text{SBI}_1}(x) + (1 - \sqrt{\mu}) \nu_B \, p_B(x) \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

Semi-parametric NSBI

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i)$$

?

Event rates estimated from simulations

Comes from theory model chosen to interpret data

j runs over different physics process (Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Example use case

$$p_{\text{ggF}}(x|\mu) = \frac{1}{\nu_{\text{ggF}}(\mu)} \left[(\mu - \sqrt{\mu}) \, \nu_S \, p_S(x) + \sqrt{\mu} \, \nu_{\text{SBI}_1} \, p_{\text{SBI}_1}(x) + (1 - \sqrt{\mu}) \nu_B \, p_B(x) \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

Semi-parametric NSBI

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i) \quad ? \quad \Rightarrow \quad \frac{p(x_i|\mu)}{p_{\mathrm{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \frac{p_j(x_i)}{p_{\mathrm{ref}}(x_i)}$$

Reference hypothesis

j runs over different physics process (Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Event rates estimated from simulations

Comes from theory model chosen to interpret data

Example use case

$$p_{\mathrm{ggF}}(x|\mu) = \frac{1}{\nu_{\mathrm{ggF}}(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S \, p_S(x) + \sqrt{\mu}\, \nu_{\mathrm{SBI}_1} \, p_{\mathrm{SBI}_1}(x) + (1 - \sqrt{\mu}) \nu_B \, p_B(x) \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

Semi-parametric NSBI

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i) \quad \Rightarrow \quad \frac{p(x_i|\mu)}{p_{\mathrm{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, \frac{p_j(x_i)}{p_{\mathrm{ref}}(x_i)}$$

?

Reference hypothesis

j runs over different physics process (Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Event rates estimated from simulations

Comes from theory model chosen to interpret data

Example use case

$$p_{\mathrm{ggF}}(x|\mu) = \frac{1}{\nu_{\mathrm{ggF}}(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S \, p_S(x) + \sqrt{\mu}\, \nu_{\mathrm{SBI}_1} \, p_{\mathrm{SBI}_1}(x) + (1 - \sqrt{\mu}) \nu_B \, p_B(x) \right]$$

$$\Rightarrow \quad \frac{p(x|\mu)}{p_S(x)} = \frac{1}{\nu(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S + \sqrt{\mu}\, \nu_{\mathrm{SBI}_1} \frac{p_{\mathrm{SBI}_1}(x)}{p_S(x)} + (1 - \sqrt{\mu}) \nu_B \frac{p_B(x)}{p_S(x)} \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

Semi-parametric NSBI

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i)$$

?

Estimated using an ensemble of networks

$$\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}$$

Reference hypothesis

j runs over different physics process (Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Event rates estimated from simulations

Comes from theory model chosen to interpret data

Example use case

$$p_{\text{ggF}}(x|\mu) = \frac{1}{\nu_{\text{ggF}}(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S \, p_S(x) + \sqrt{\mu}\, \nu_{\text{SBI}_1} \, p_{\text{SBI}_1}(x) + (1 - \sqrt{\mu}) \nu_B \, p_B(x) \right]$$

$$\frac{p(x|\mu)}{p_S(x)} = \frac{1}{\nu(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S + \sqrt{\mu}\, \nu_{\text{SBI}_1} \frac{p_{\text{SBI}_1}(x)}{p_S(x)} + (1 - \sqrt{\mu}) \nu_B \frac{p_B(x)}{p_S(x)} \right]$$

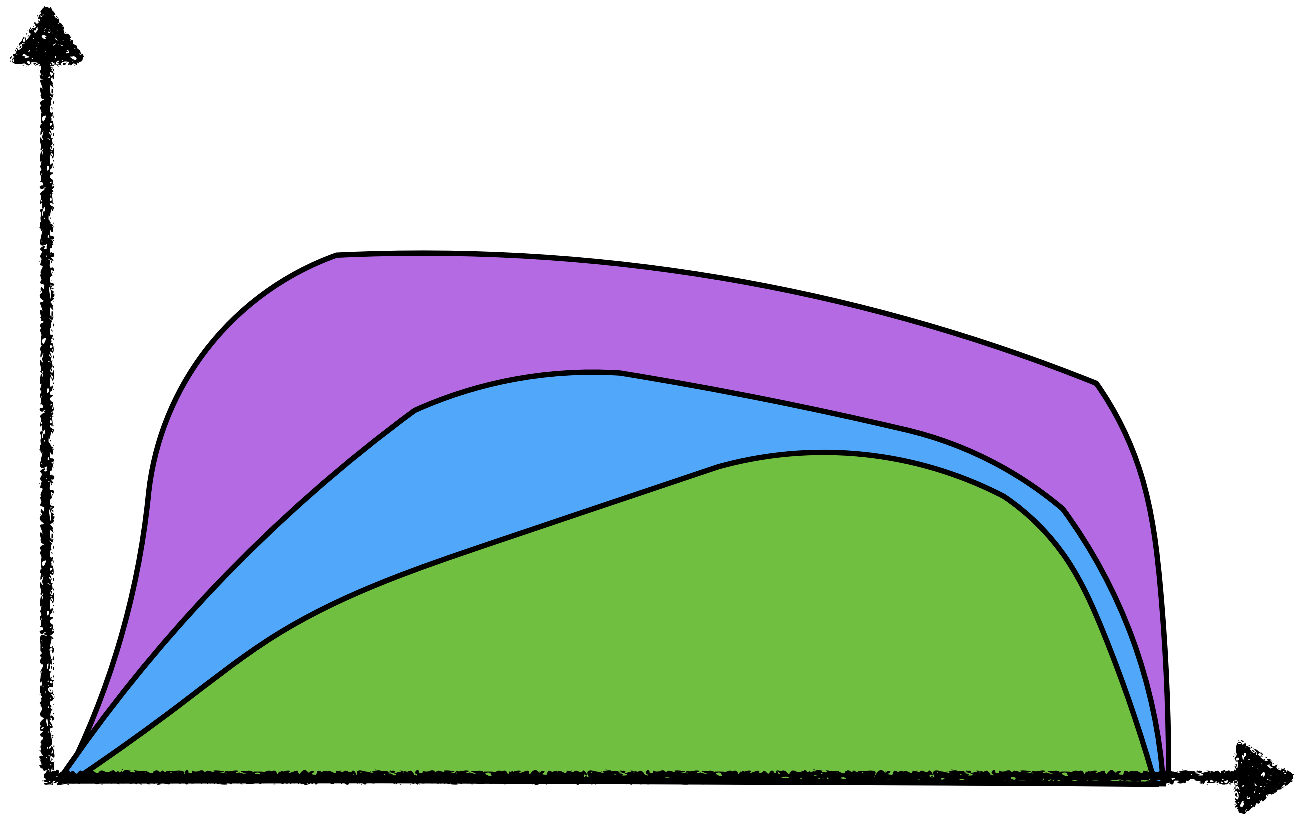
$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

Robust, parameterised classifier without parameterising

H_{ref} : Reference hypothesis

H_μ

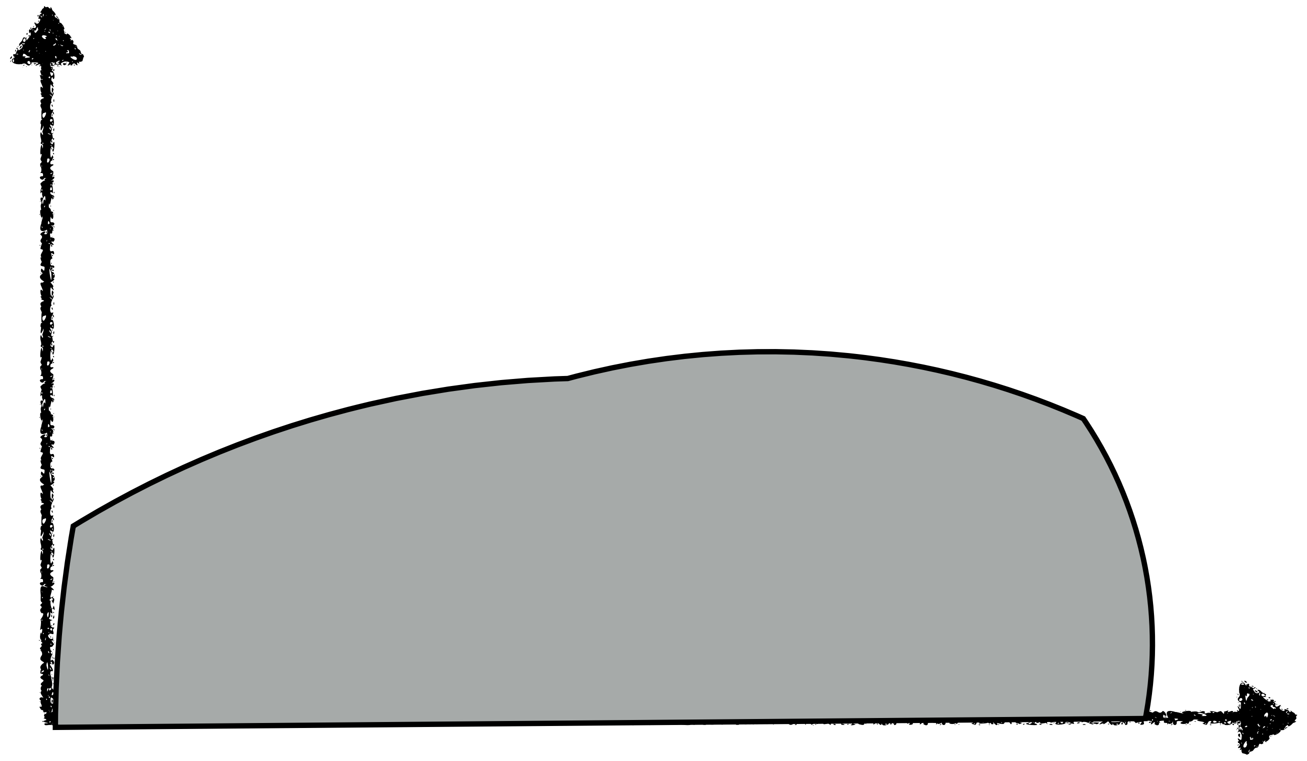
Hypothesis



VS

H_{ref}

Reference hypothesis



$$\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \; \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}$$

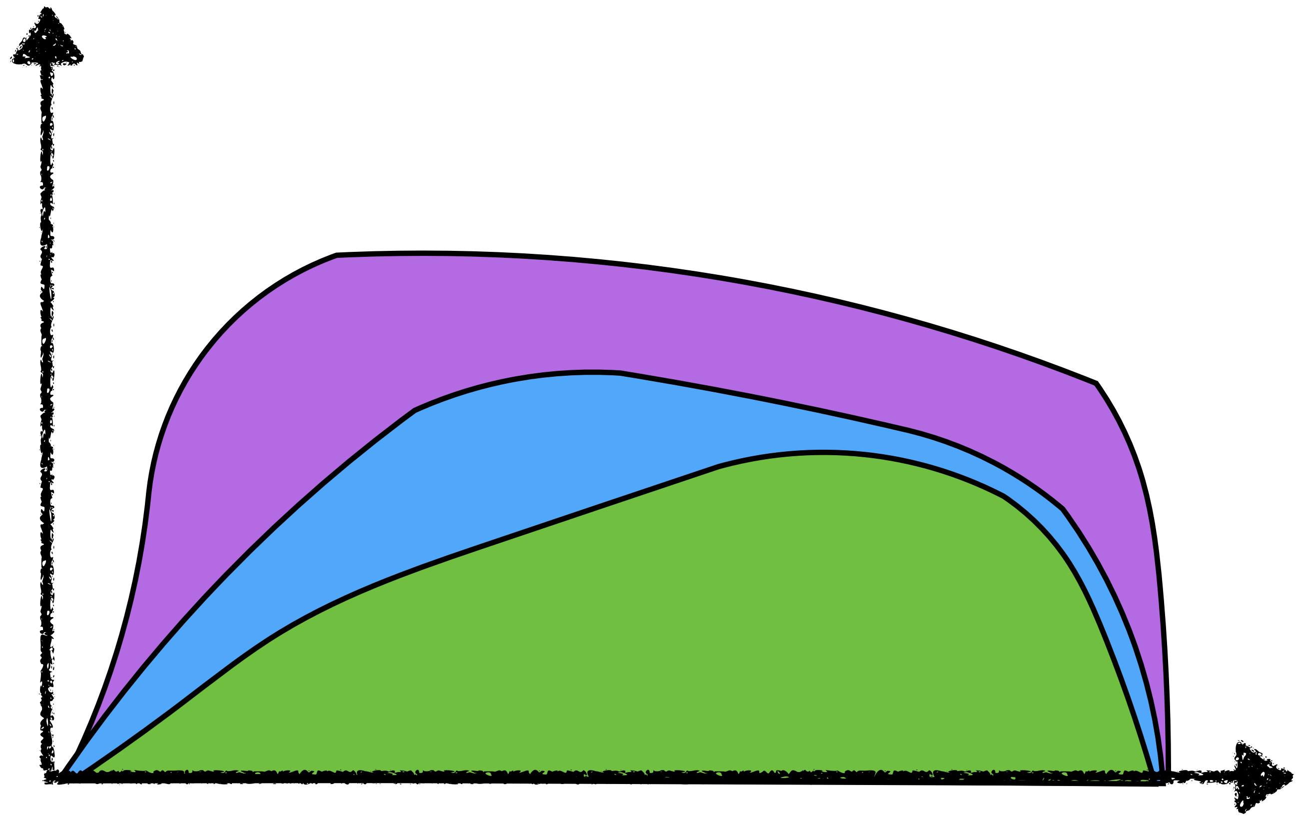
A separate classifier per physics process j
(Eg. $gg \rightarrow H^* \rightarrow 4l$, $gg \rightarrow ZZ \rightarrow 4l$)

Robust, parameterised classifier without parameterising

H_{ref} : Reference hypothesis

H_μ

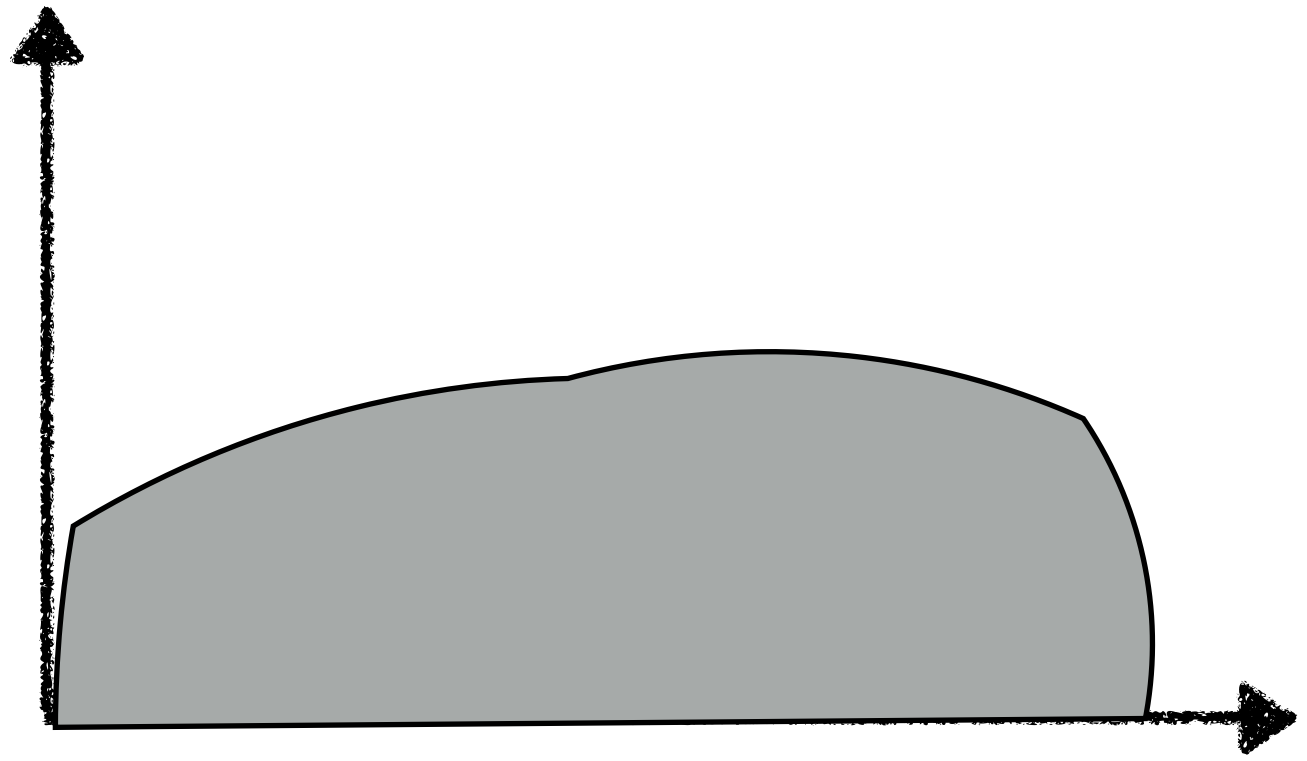
Hypothesis



VS

H_{ref}

Reference hypothesis

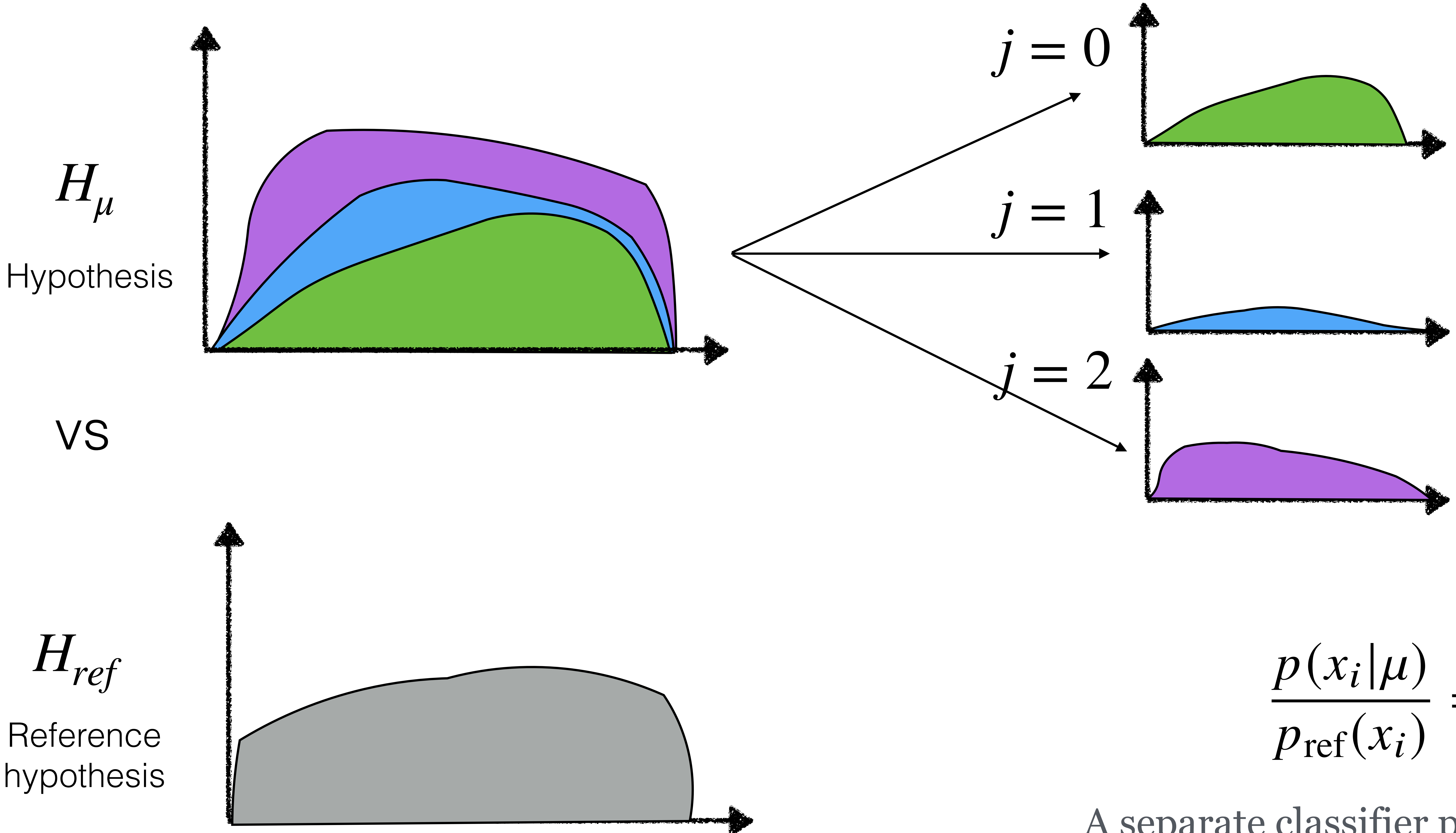


$$\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}$$

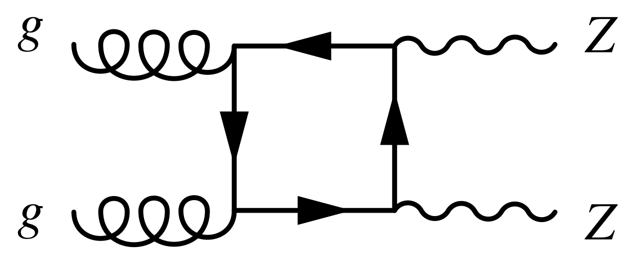
A separate classifier per physics process j
(Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Robust, parameterised classifier without parameterising

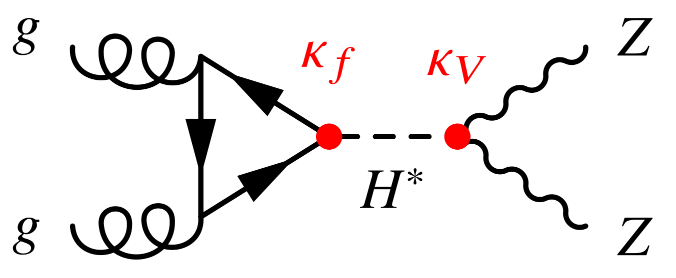
H_{ref} : Reference hypothesis



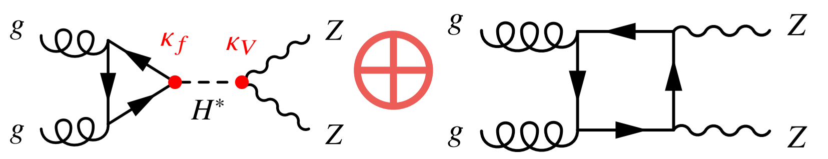
$f_0(\mu) = 1$



$f_1(\mu) = \mu$



$f_2(\mu) = \sqrt{\mu}$



$$\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}$$

A separate classifier per physics process j
(Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Robust, parameterised classifier without parameterising

H_{ref} : Reference hypothesis

$$\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}$$

Analytically parameterised in μ, allows to get LR for any hypothesis μ without training parameterised networks !

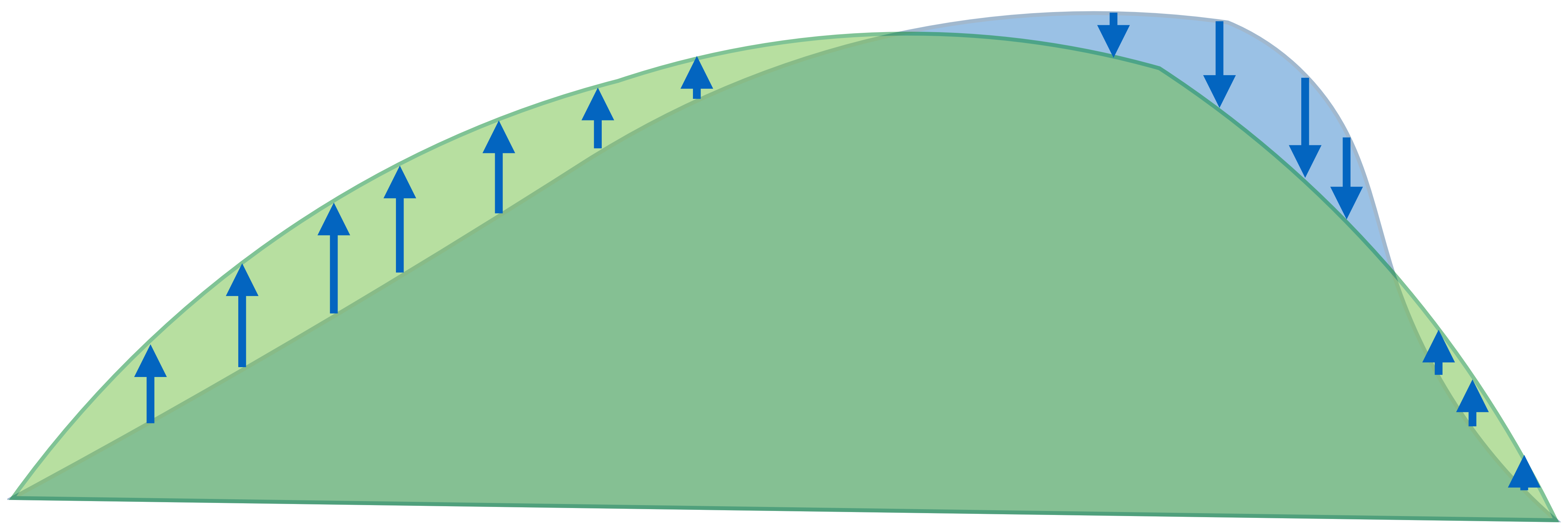
A separate classifier per physics process j
(Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Is this dream attainable in practice?

- How to ensure robustness? By design and validation
- How to incorporate systematic uncertainties?
- How to perform Neyman construction in high dimensions?

Validate quality of LR estimation with re-weighting task

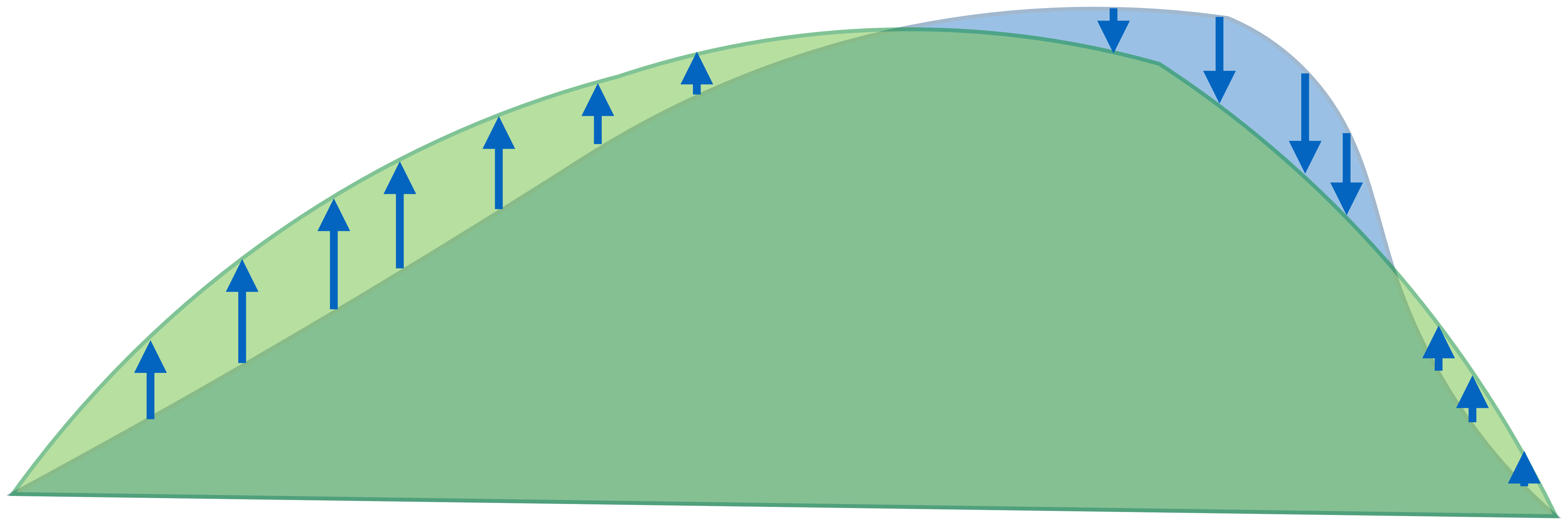
Reweighting: Calculate weights w_i for events x_i in blue sample to match green sample



Validate quality of LR estimation with re-weighting task

Reweighting: Calculate weights w_i for events x_i in blue sample to match green sample

$$w_i = r(x_i, \mu_0, \mu_1) = \frac{p(x_i \mid \mu_0)}{p(x_i \mid \mu_1)}$$

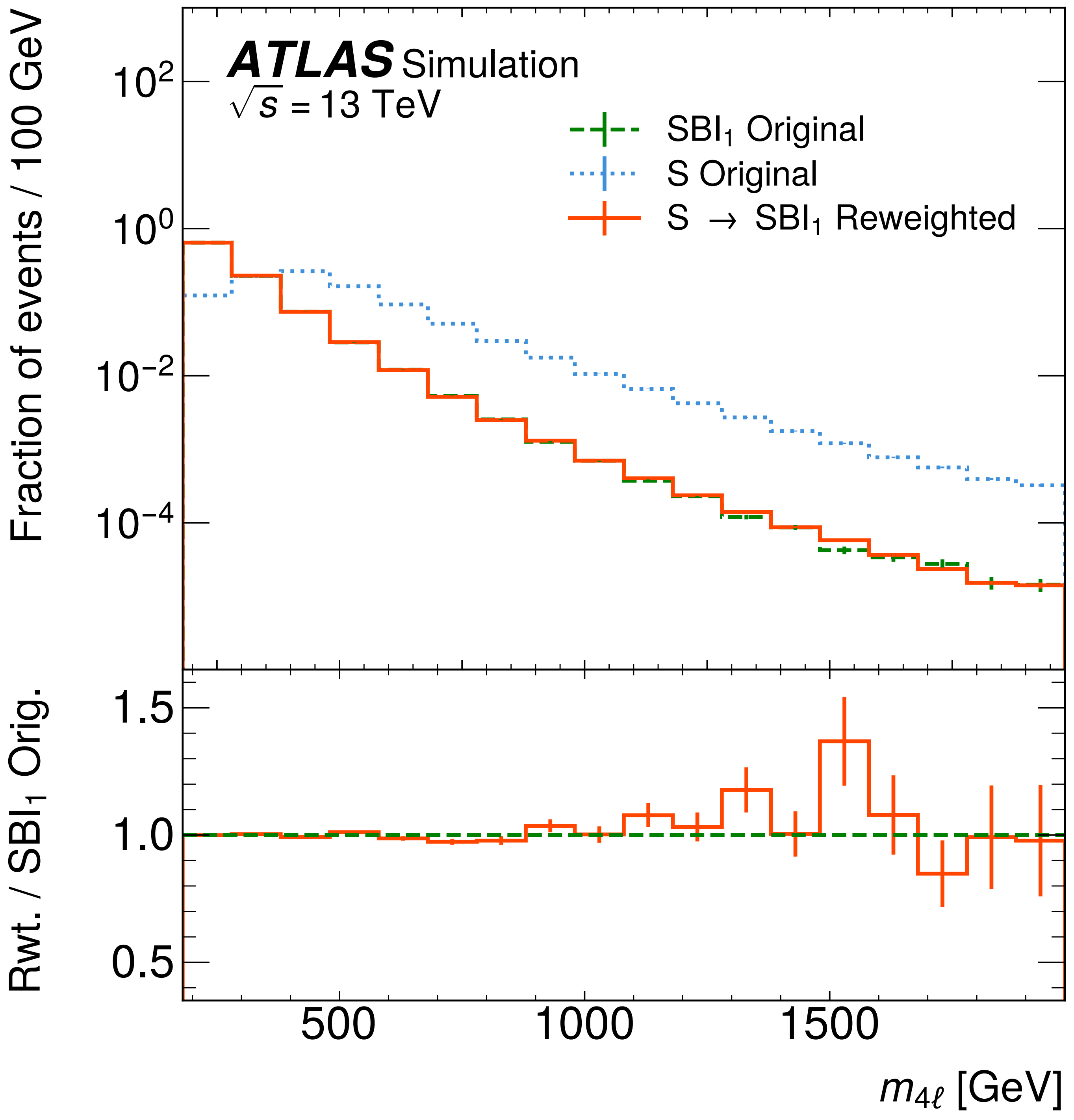


Probability ratios already estimated using an ensemble of networks

Re-weight closures

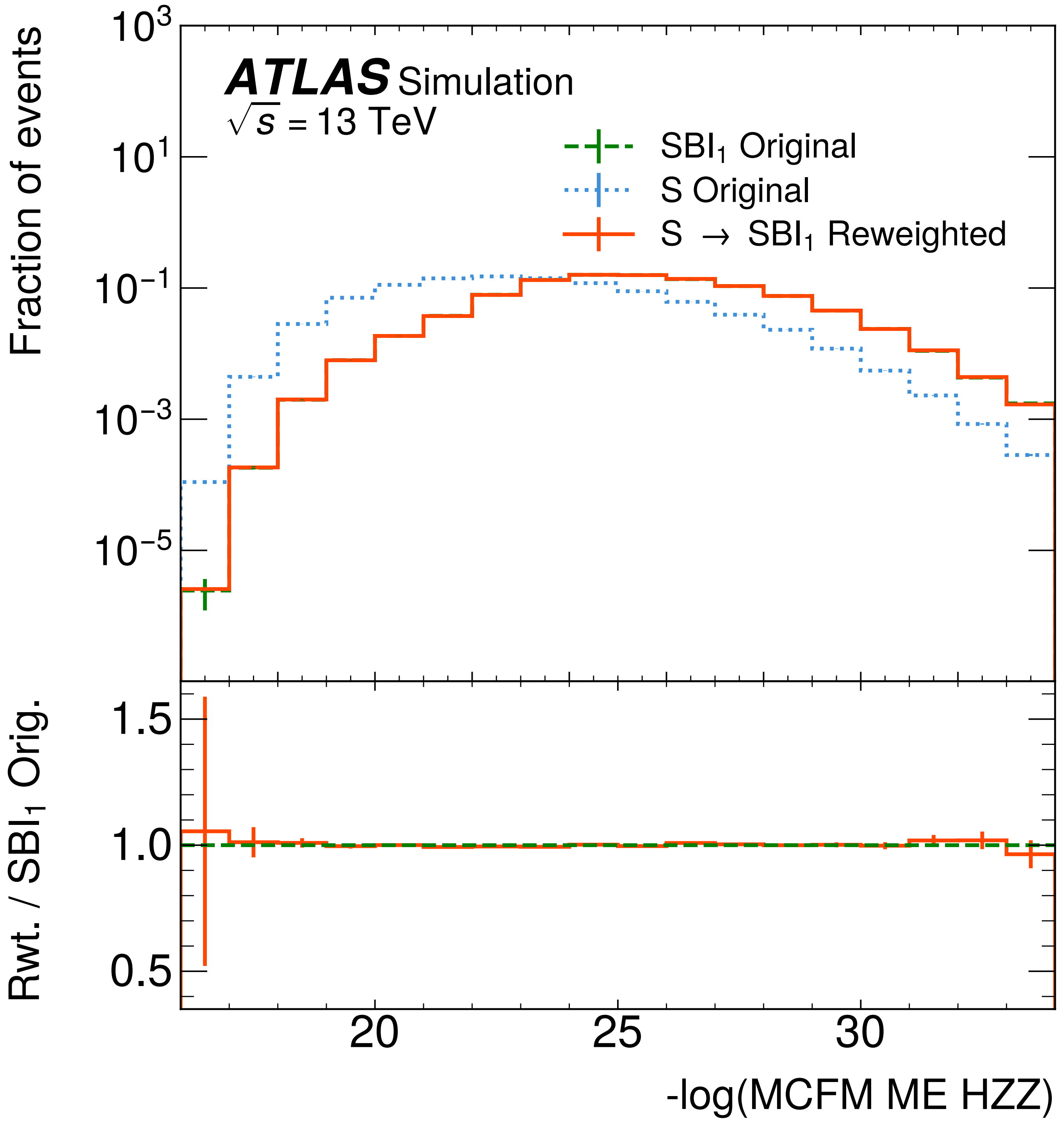
Source
Target
RW

Variable used in training



m_{4l}

High-level variable
never used in training



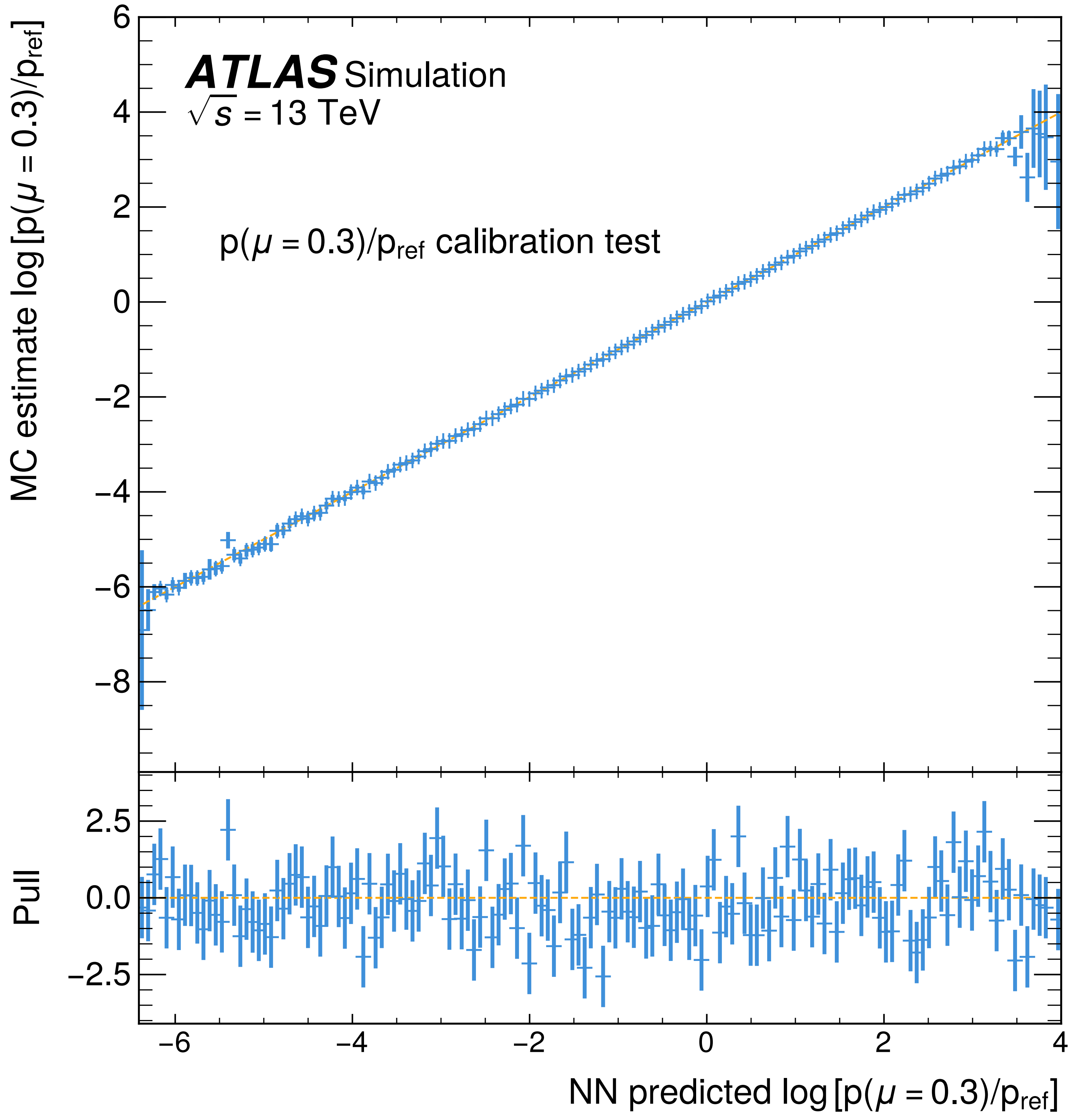
Matrix-Element-based Observable
(ggF from MCFM)

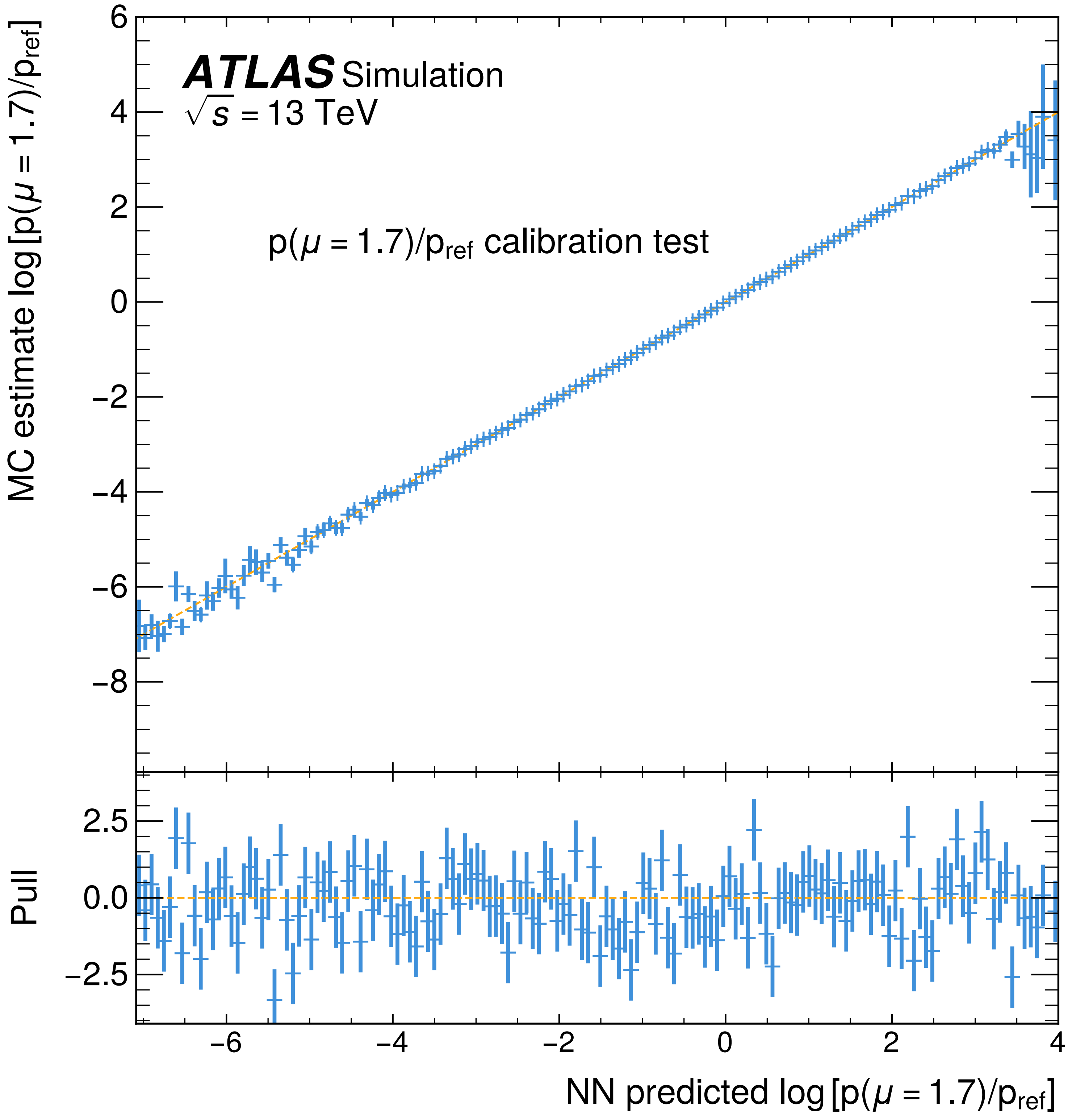
Calibration curves of probability density ratios

$$\frac{p_{\mu=0.3}(x_i)}{p_{ref}(x_i)}$$

$$\frac{p_{\mu=1.7}(x_i)}{p_{ref}(x_i)}$$

Binned estimate





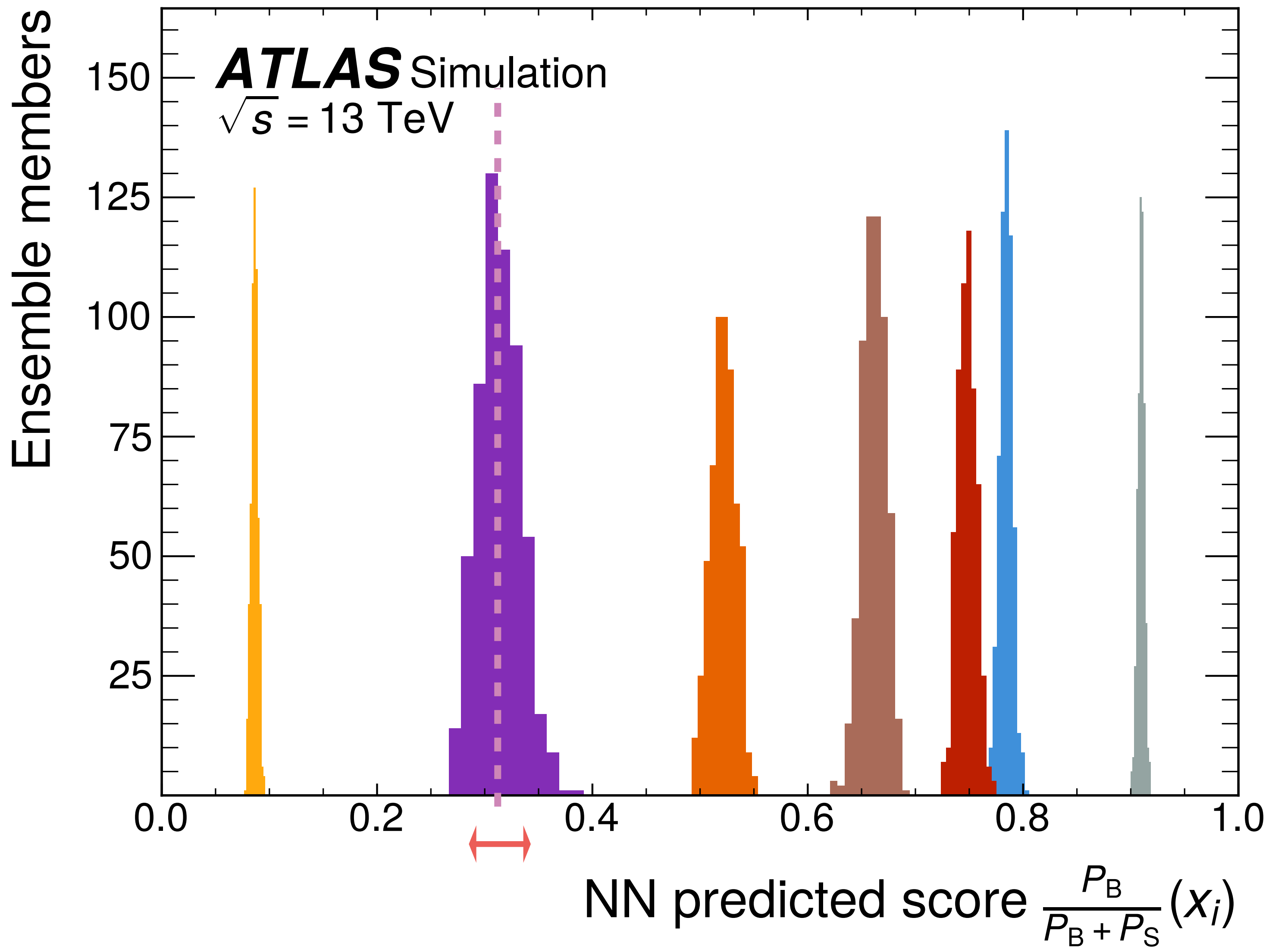
Perfect calibration would give $y = x$

Ensemble prediction

Ensemble prediction

Uncertainties on the network estimates themselves

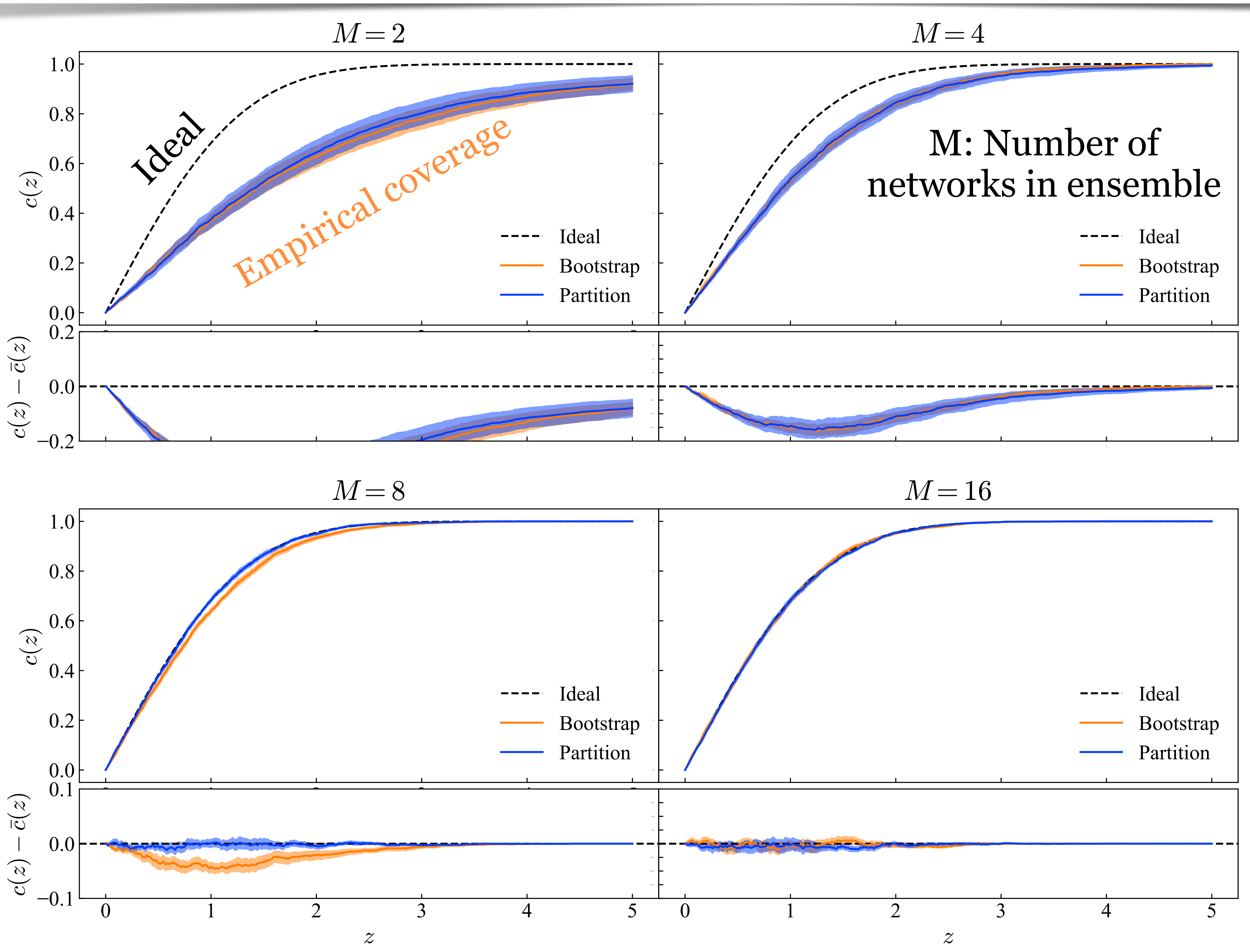
Distribution of NN predictions for example events



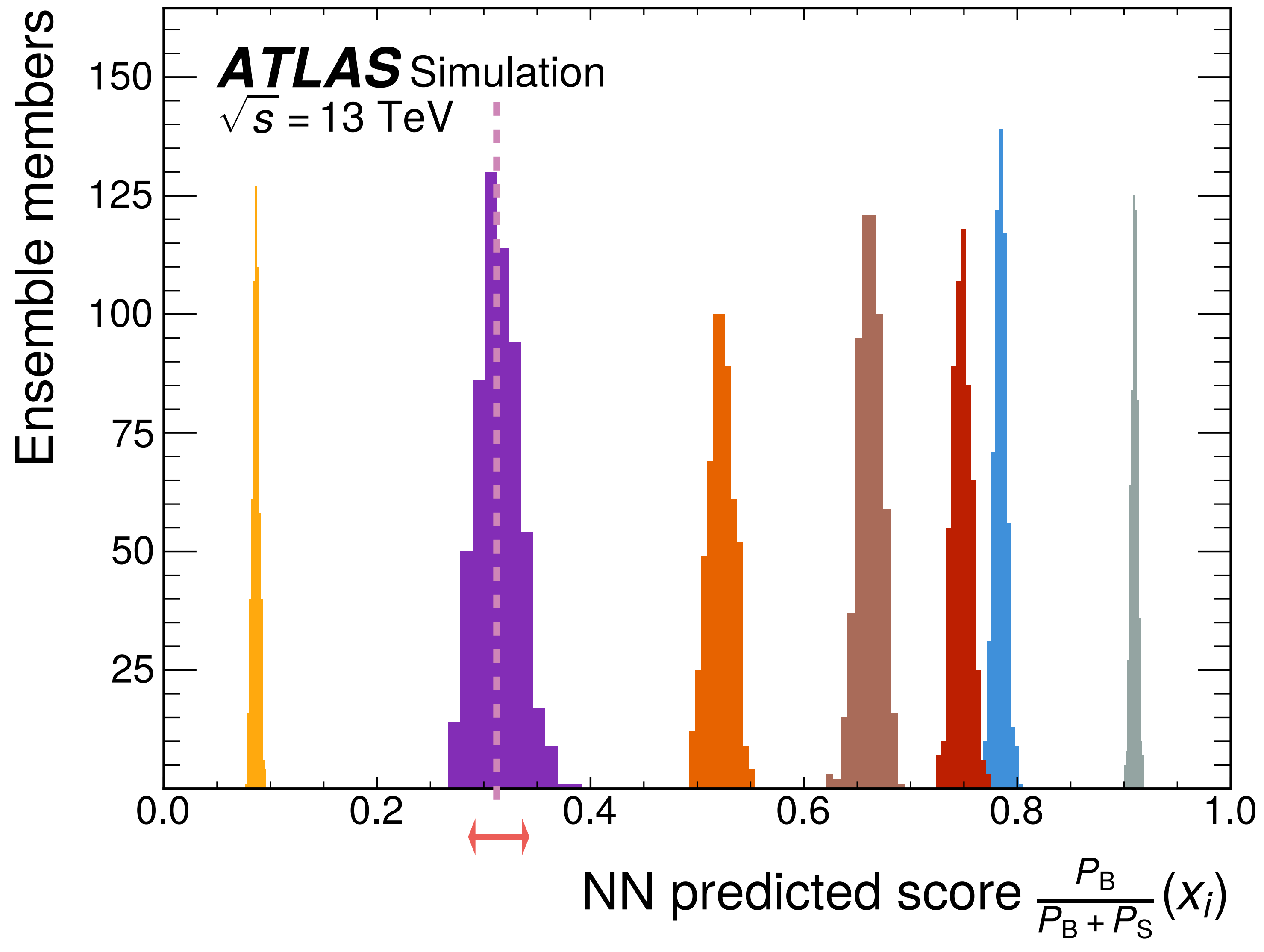
Uncertainties on the network estimates themselves

Frequentist Uncertainties on Neural Density Ratios with $w_i f_i$ Ensembles

Sean Benevedes[1, 2, *] and Jesse Thaler[1, 2, †]
[1] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States*
[2] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*



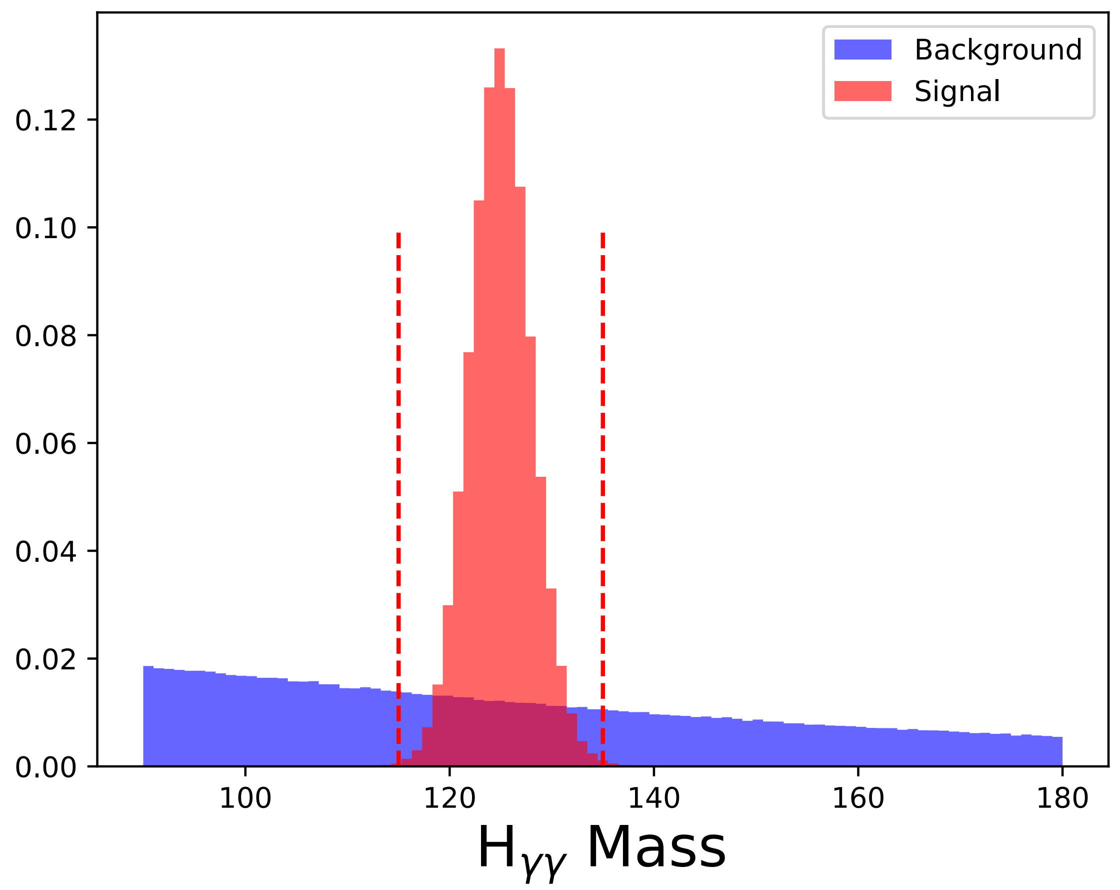
Distribution of NN predictions for example events

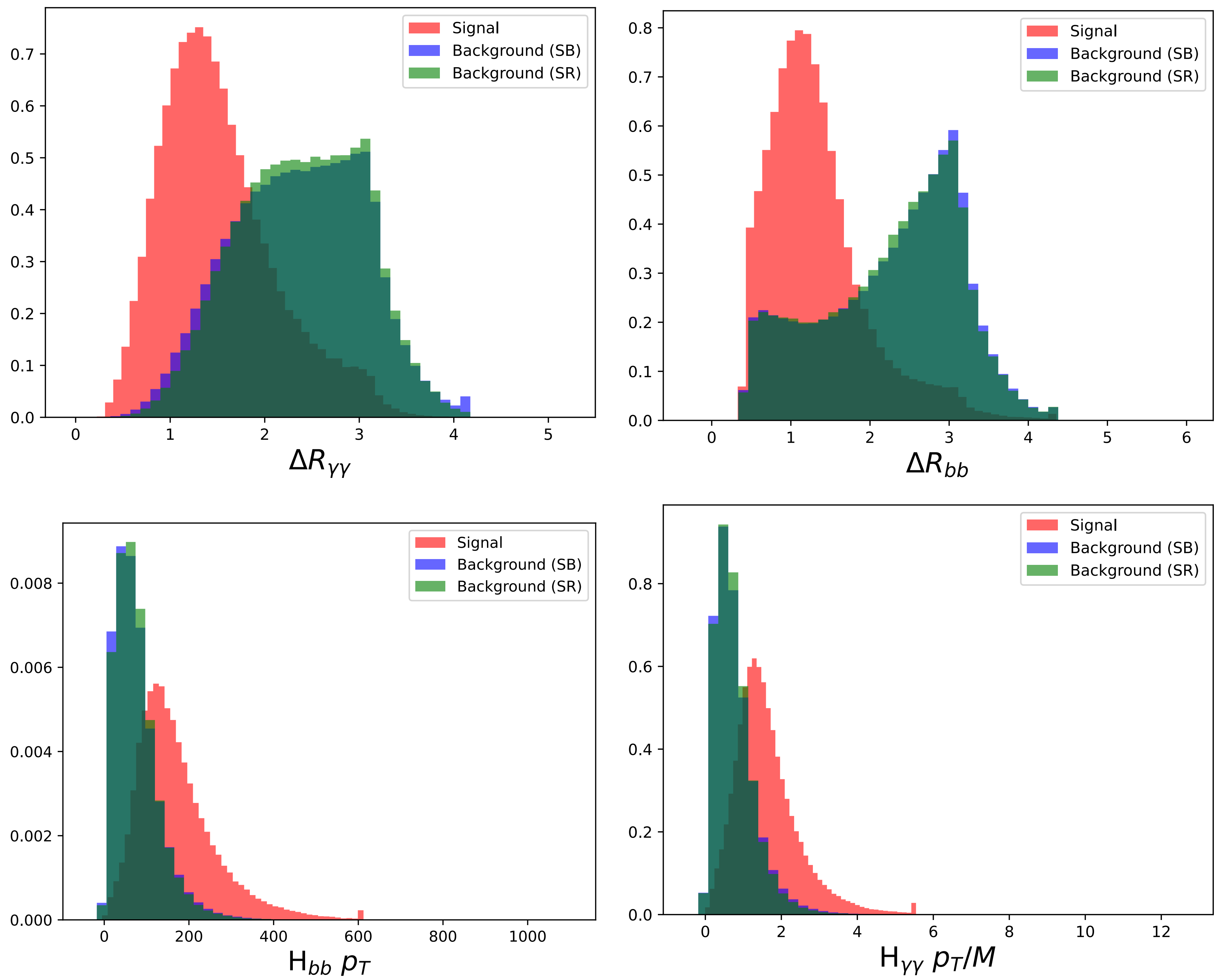


- Genuine 95% coverage, with much smaller ensembles
- Elegant, mathematically motivated method to estimate uncertainties

arXiv:2506.00113: Benevedes & Thaler

Unreliable simulator ?





- Training data can come from control regions of real data
- **Data-driven background estimation** techniques work with NSBI
- Amram & Szewc uses generative models instead of classifiers for NSBI

arXiv:2506.06438: Amram & Szewc

Is there more to simulation-based inference?

So much more! Diagnostics, Neyman construction, focusing the power of a test…
Can discuss if there is time.

But, let’s focus on the basics in today’s tutorial

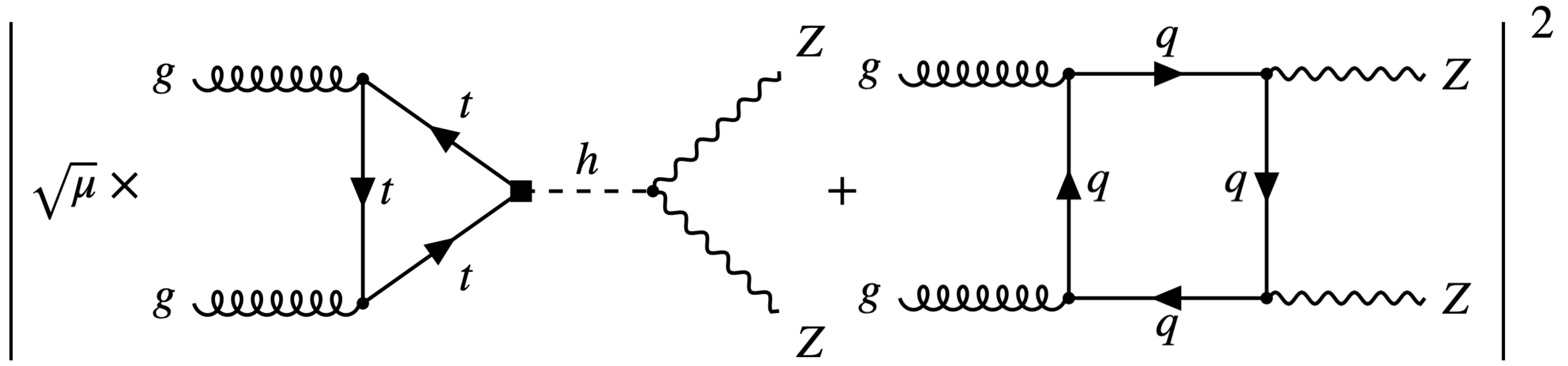
This tutorial (by Eddie McGrady)

Based on dataset and code from **Tae Hyoun Park**, originally for arXiv:2507.02032

- Simplified dataset:
  - Physics processes:
    - S: $gg \rightarrow H^* \rightarrow 4l$
    - B: $gg \rightarrow ZZ \rightarrow 4l$
    - Combined SBI
- No dominant qqZZ background, no EW processes
- No systematics

- The background sample will be used as 'reference', so $\frac{P_B}{P_{ref}} = 1$

- Only 2 density ratios to be estimated: $\frac{P_S}{P_{ref}}$ & $\frac{P_{SBI}}{P_{ref}}$

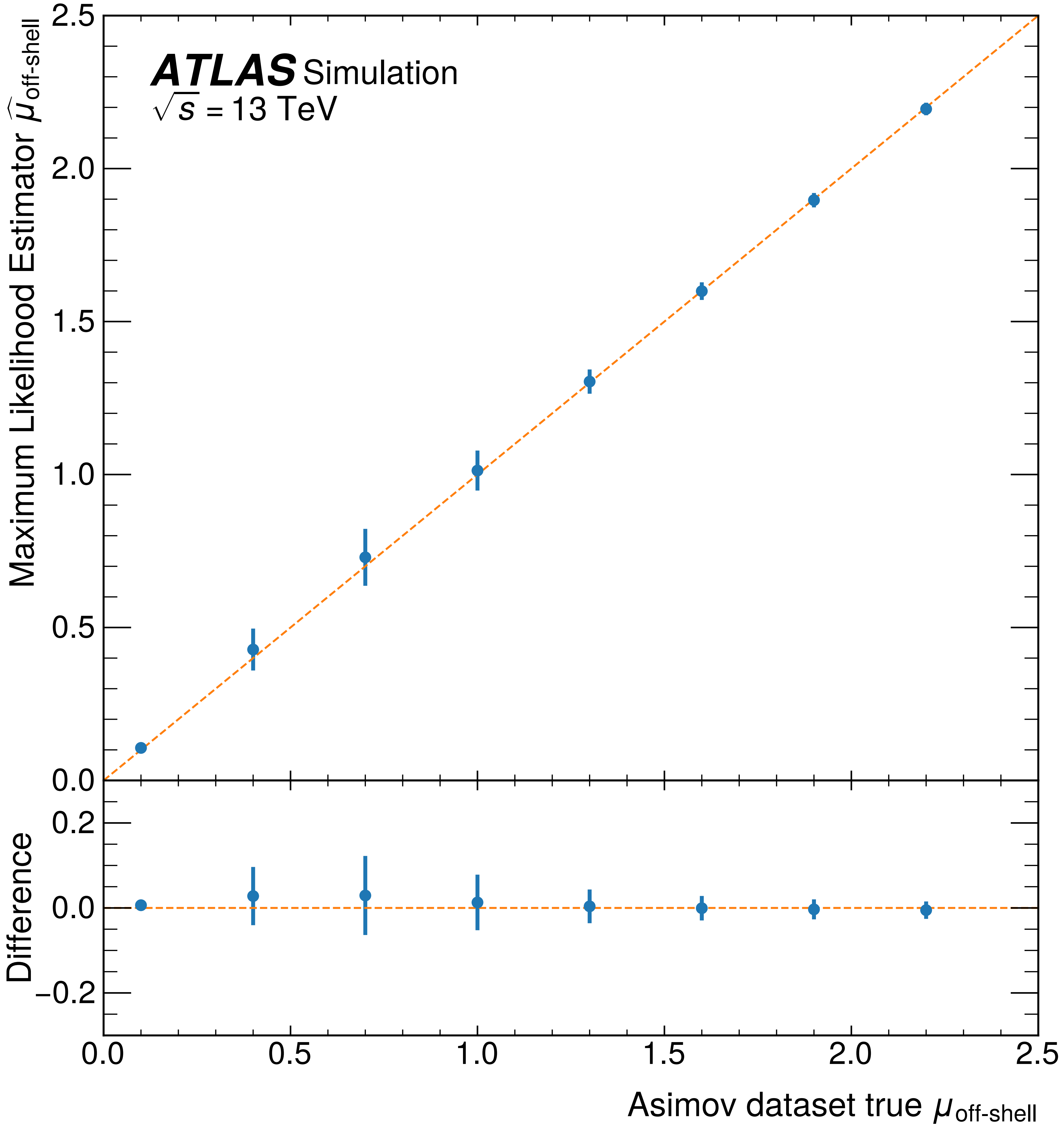


Let’s play with the notebooks !

More diagnostics

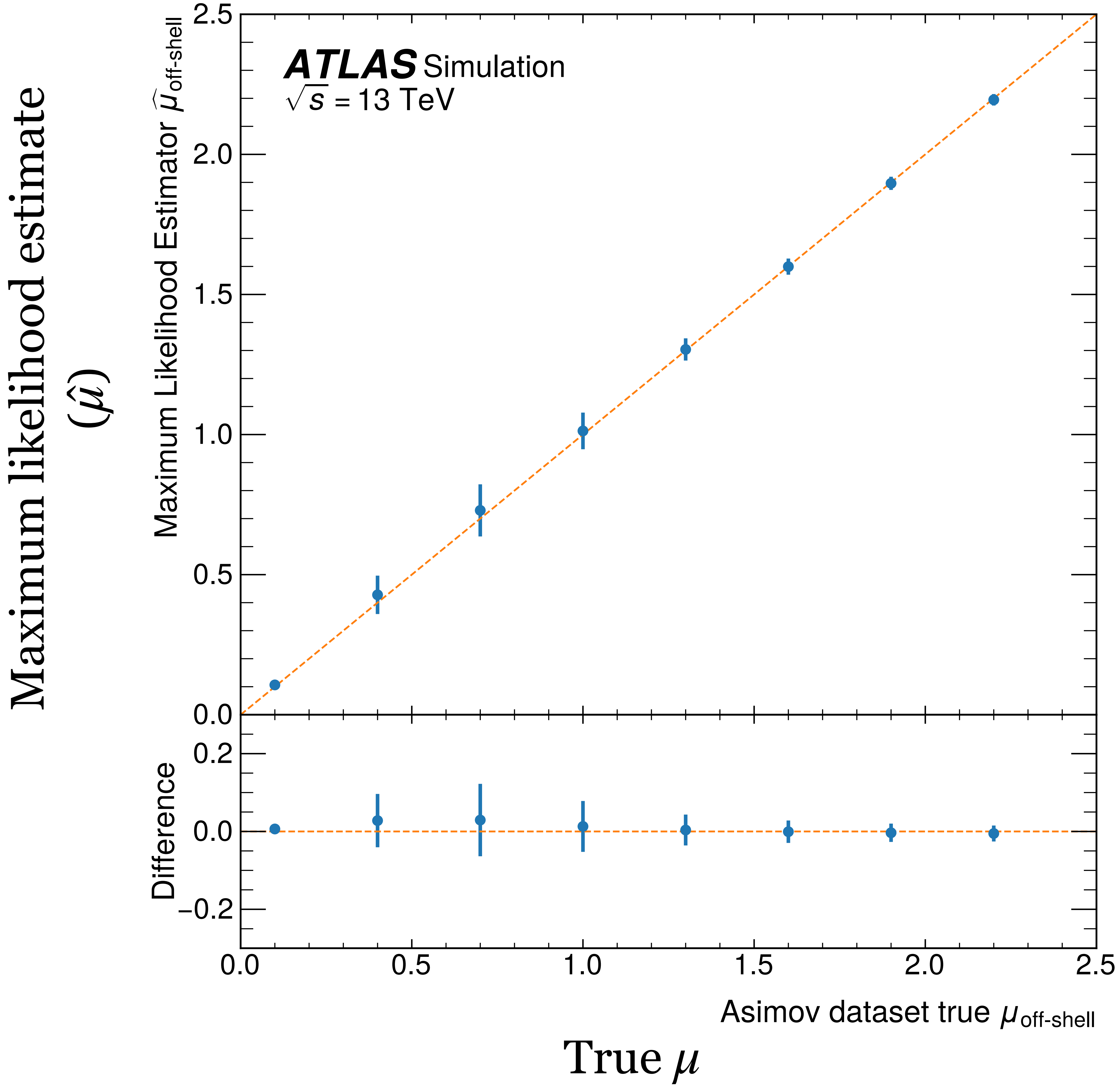
Testing full analysis on samples from different values of μ

Maximum likelihood estimate ($\hat{\mu}$)



True μ

Testing full analysis on samples from different values of μ



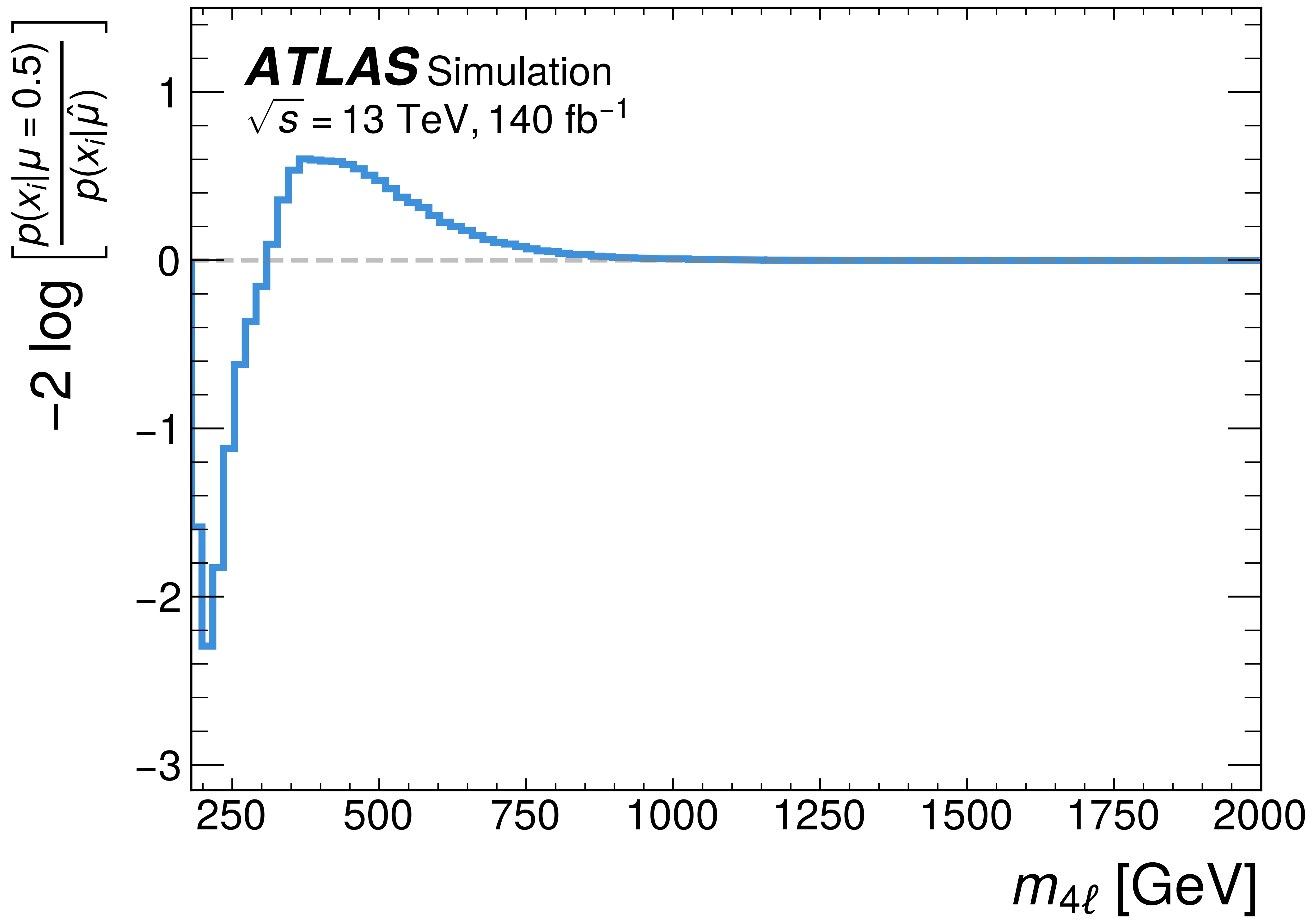
No bias: Method recovers correct value of μ on average (Correct MLE when tested on the median 'Asimov dataset')

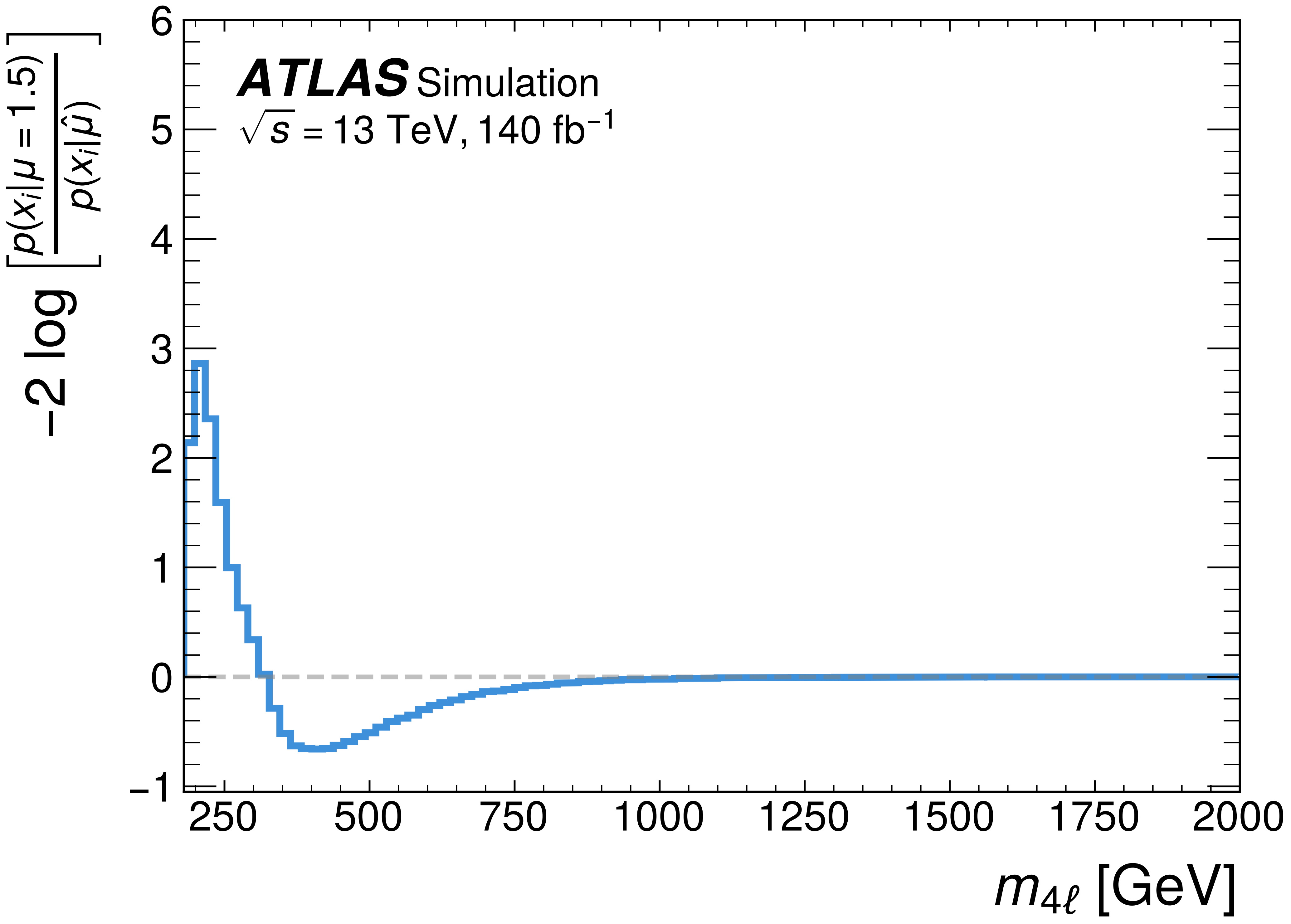
And many more diagnostics, data vs MC validation (see backup)

Interpretability:
Which phase space favours one hypothesis over another?

$$-2 \cdot log \frac{P(x_i \,|\, \mu = 0.5)}{P(x_i \,|\, \mu = 1)}$$

$$-2 \cdot log \frac{P(x_i \,|\, \mu = 1.5)}{P(x_i \,|\, \mu = 1)}$$

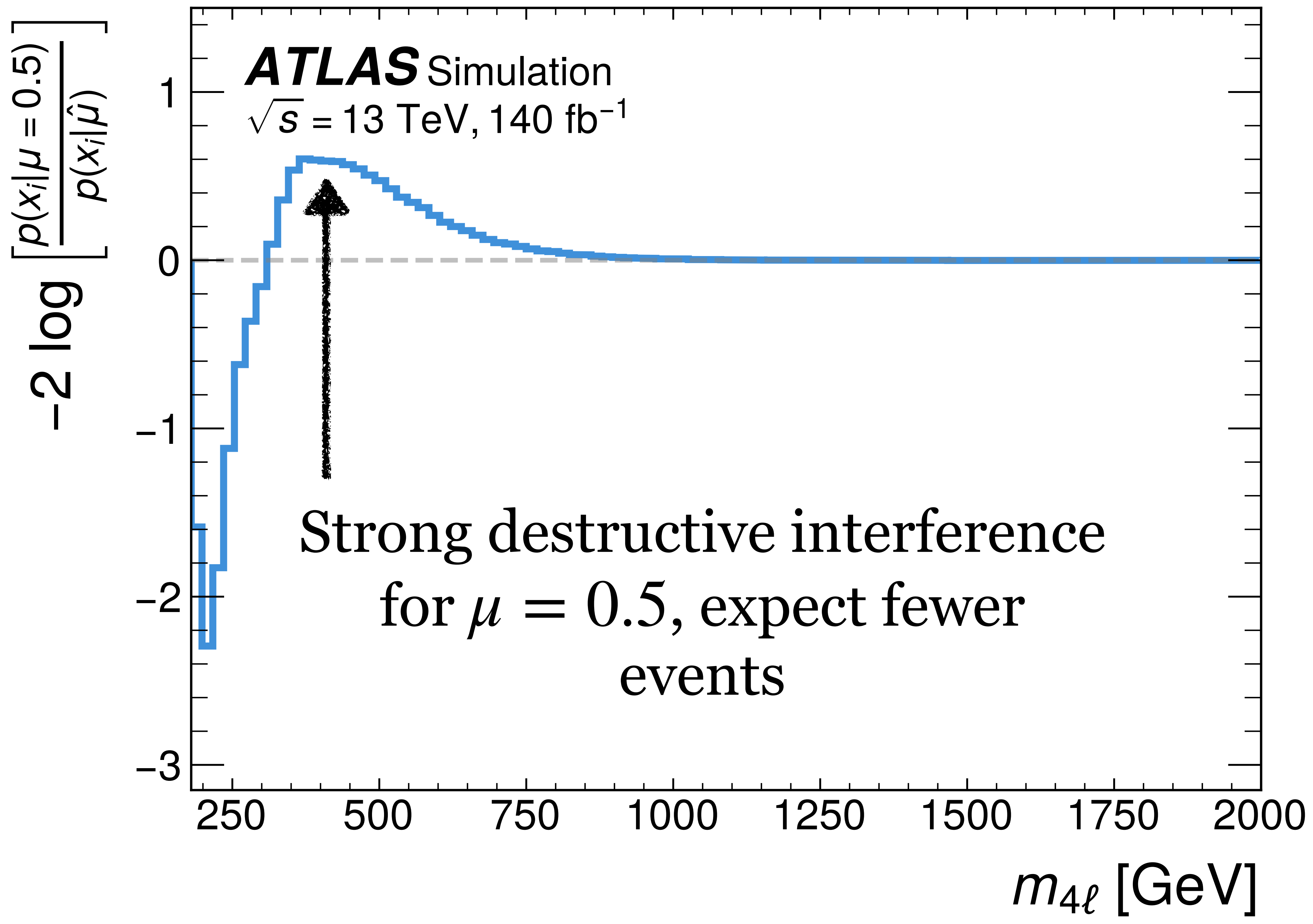


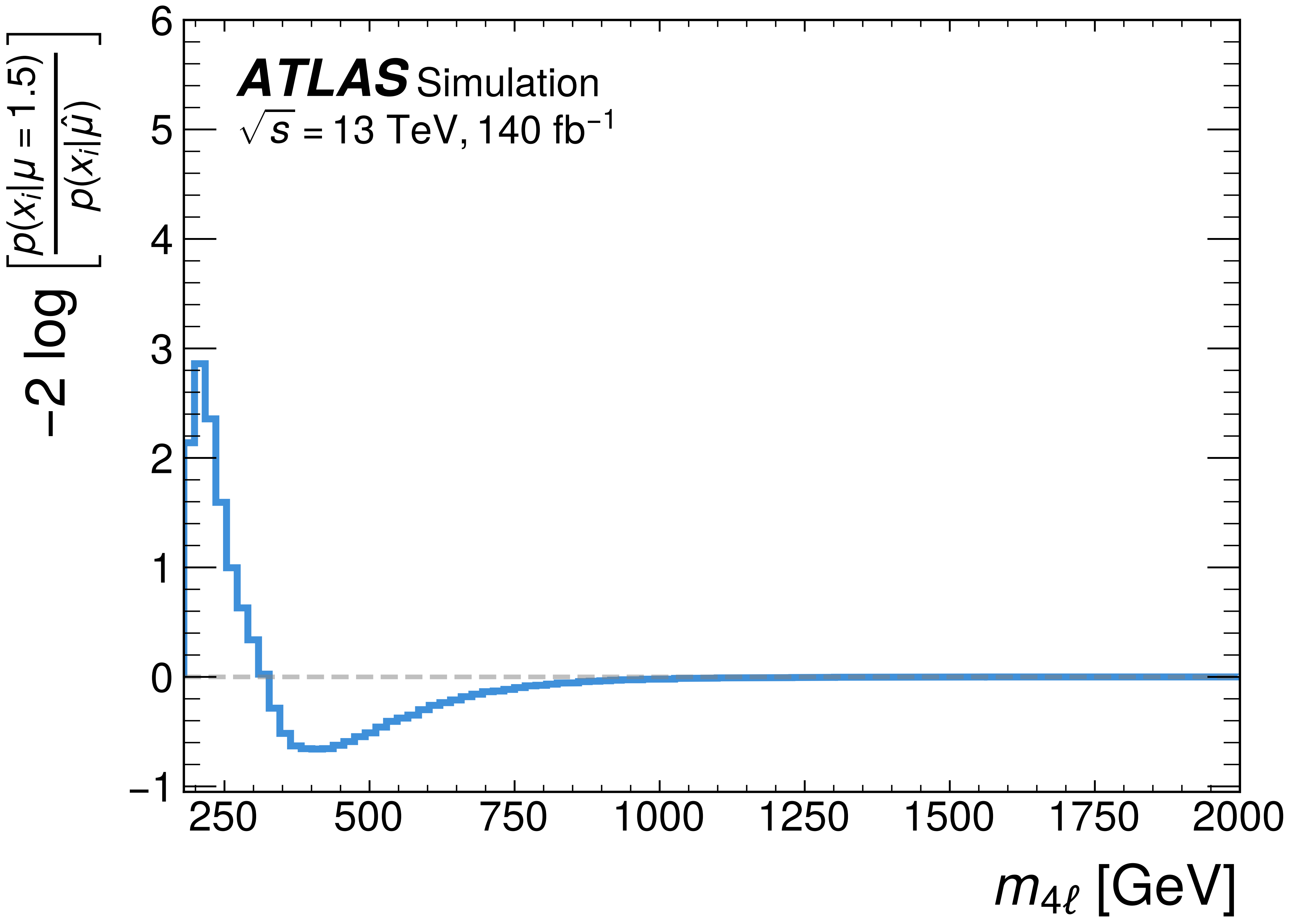


Interpretability: Which phase space favours one hypothesis over another?

$$-2 \cdot log \frac{P(x_i | \mu = 0.5)}{P(x_i | \mu = 1)}$$

$$-2 \cdot log \frac{P(x_i | \mu = 1.5)}{P(x_i | \mu = 1)}$$

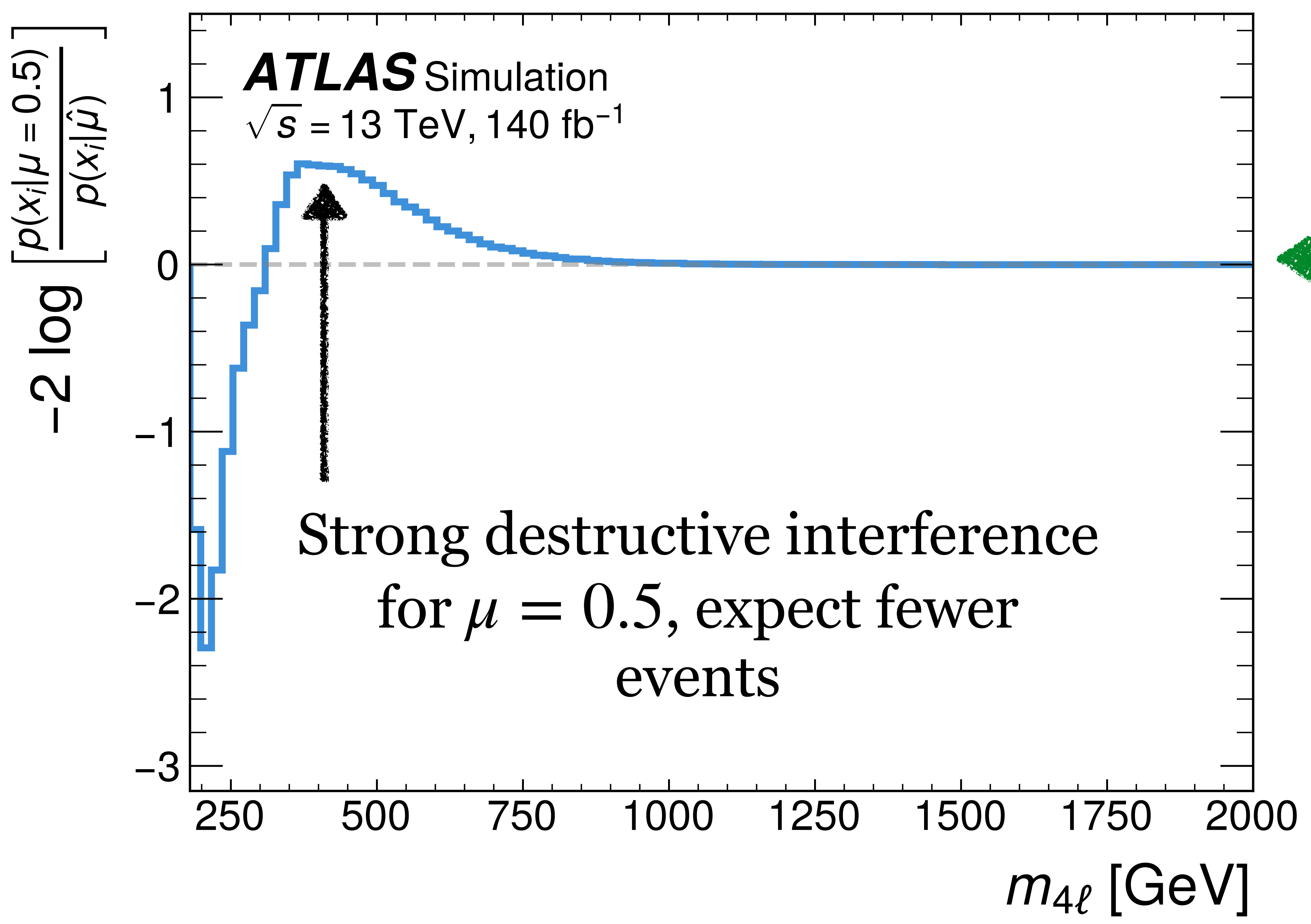


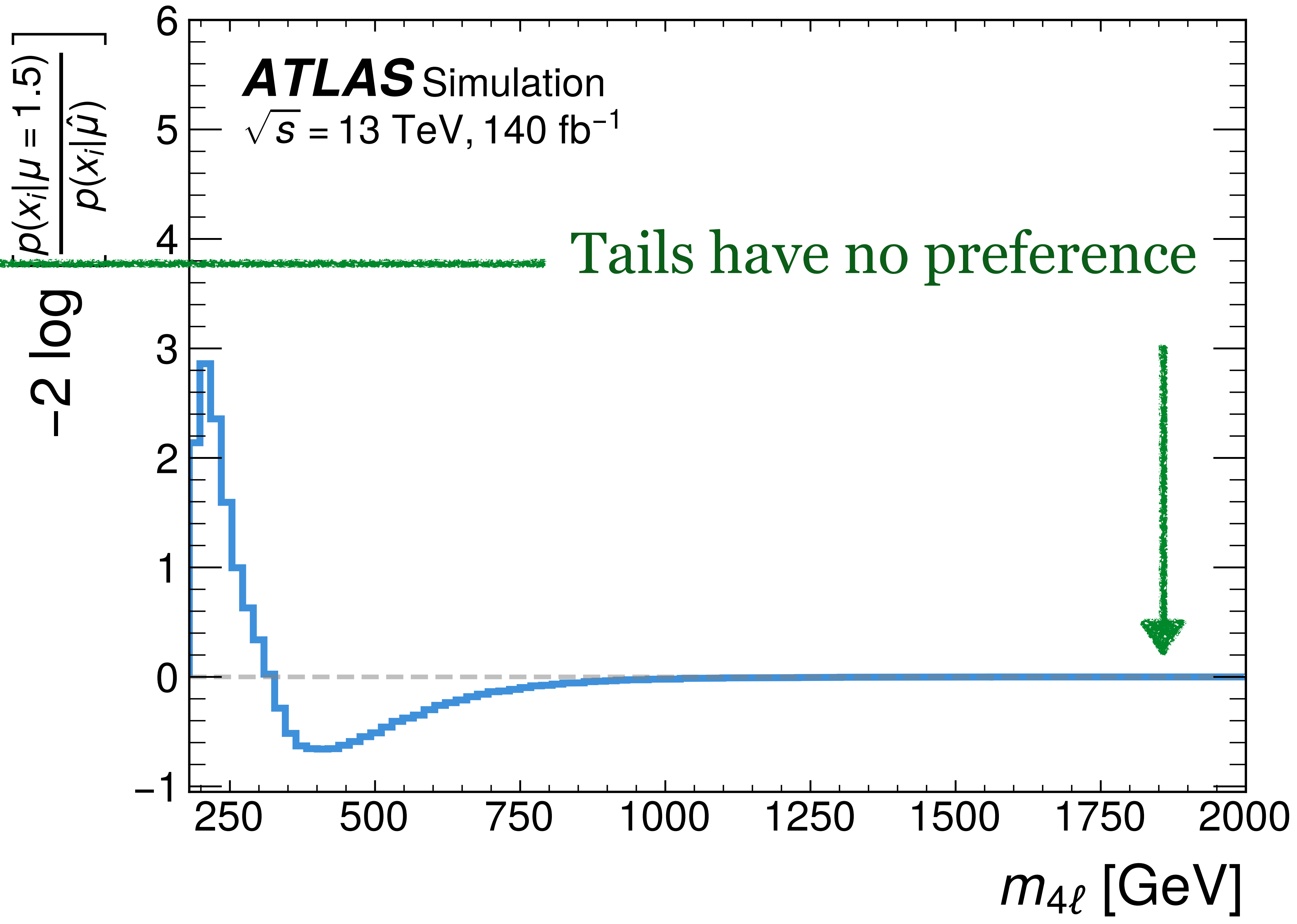


Interpretability: Which phase space favours one hypothesis over another?

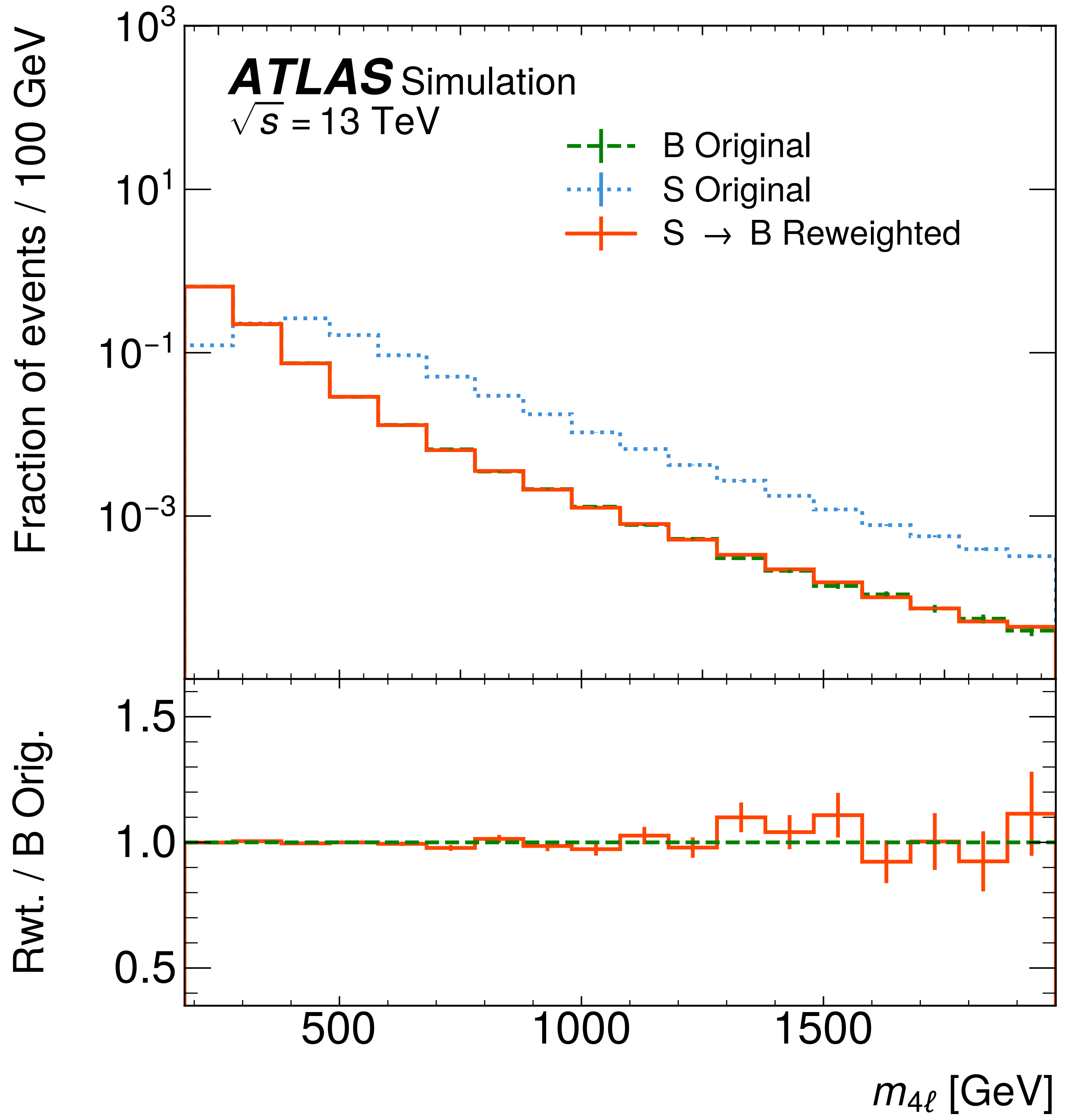
$$-2 \cdot log \frac{P(x_i | \mu = 0.5)}{P(x_i | \mu = 1)}$$

$$-2 \cdot log \frac{P(x_i | \mu = 1.5)}{P(x_i | \mu = 1)}$$

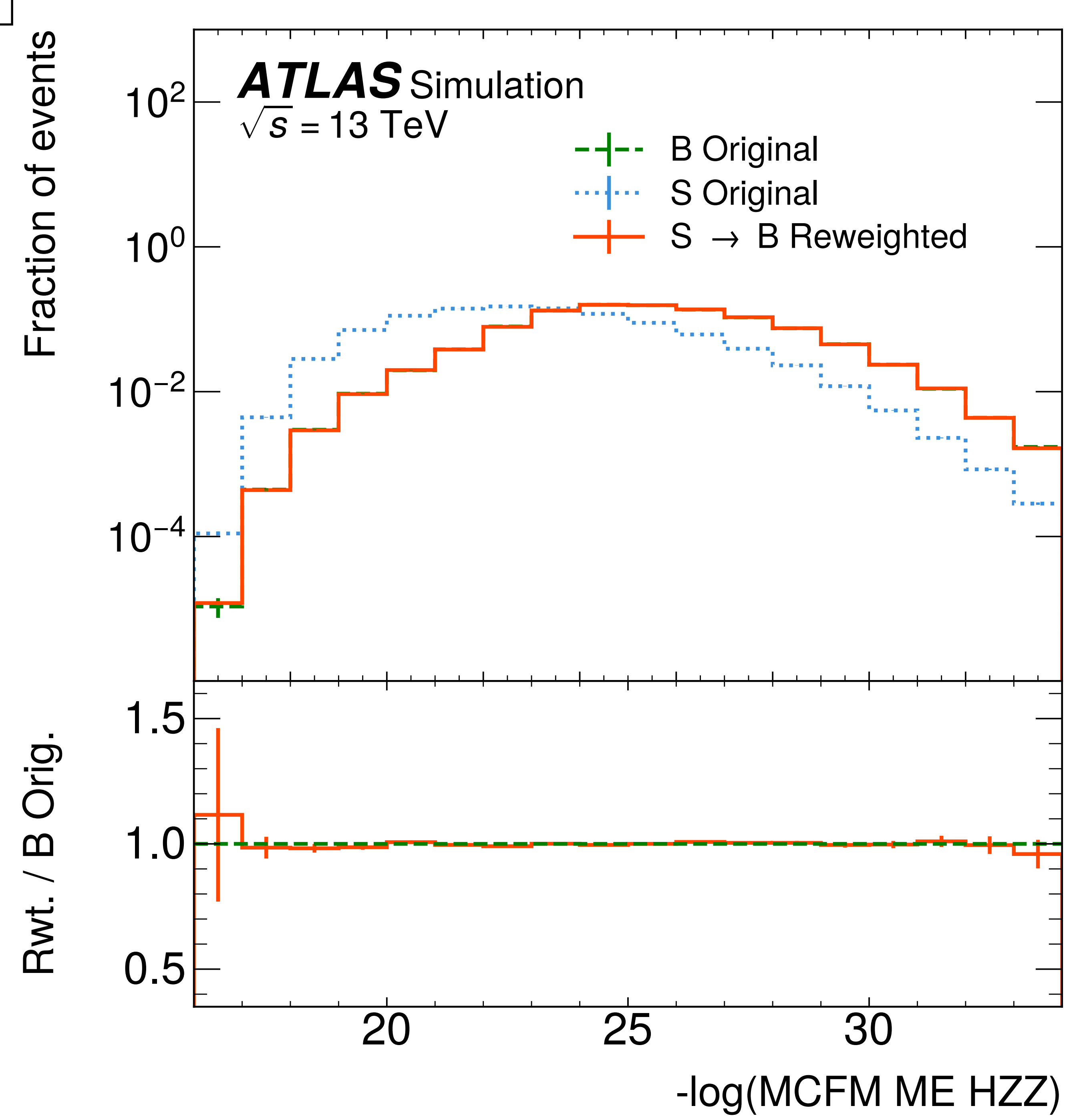




Re-weight closures for B



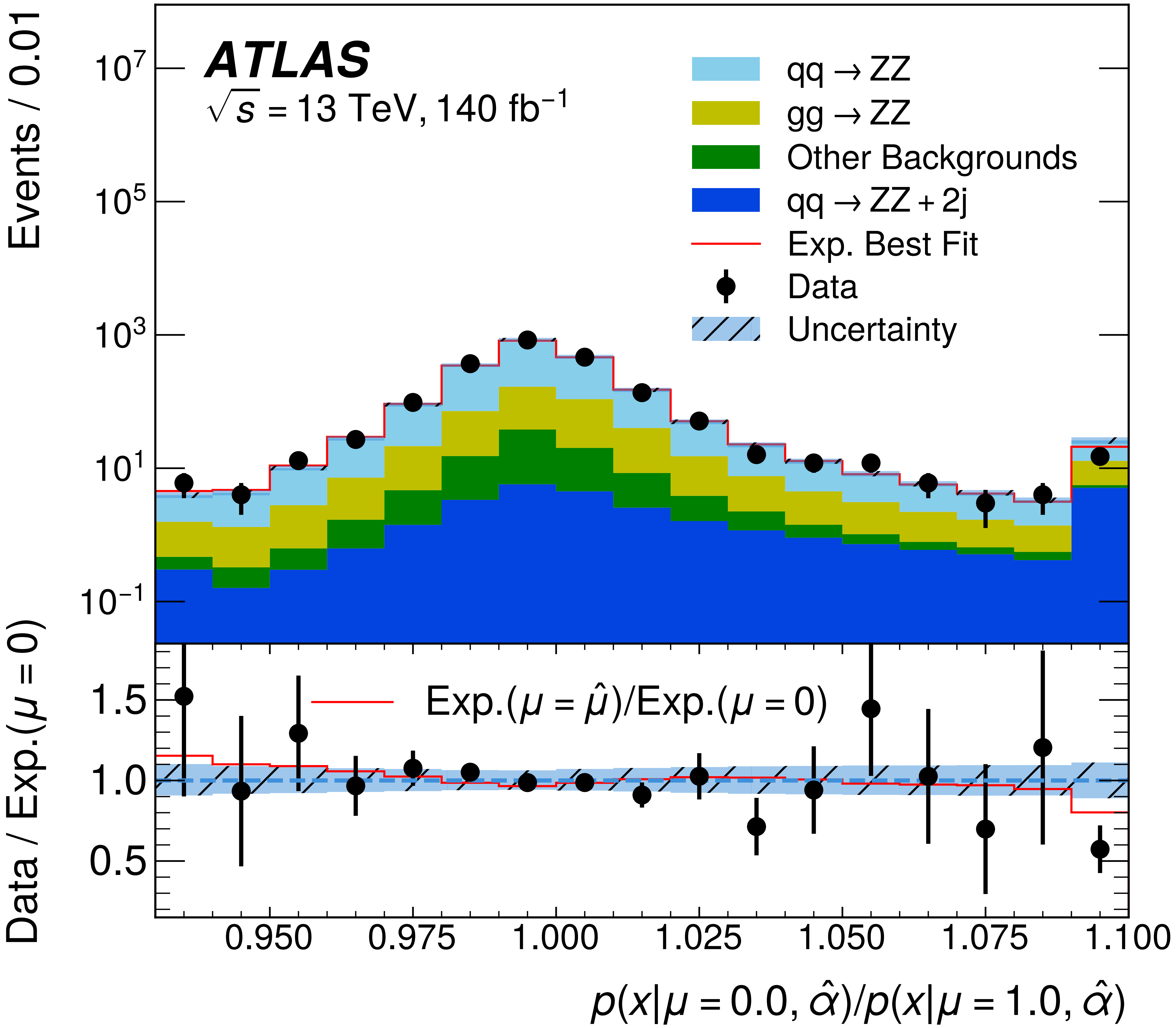
m_{4l}



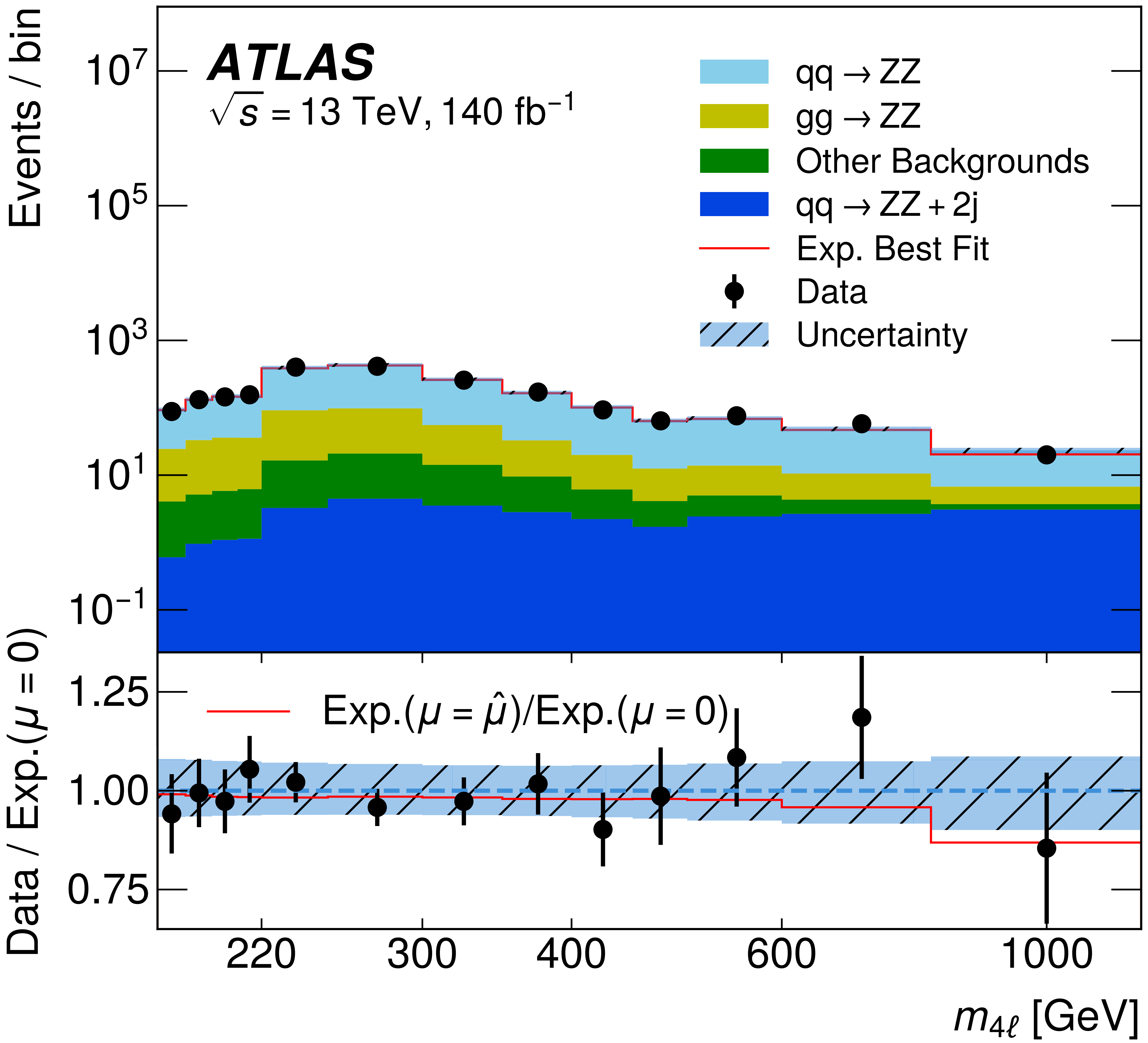
Matrix-Element-based Observable
(ggF from MCFM)

Data-MC validation

NN observable

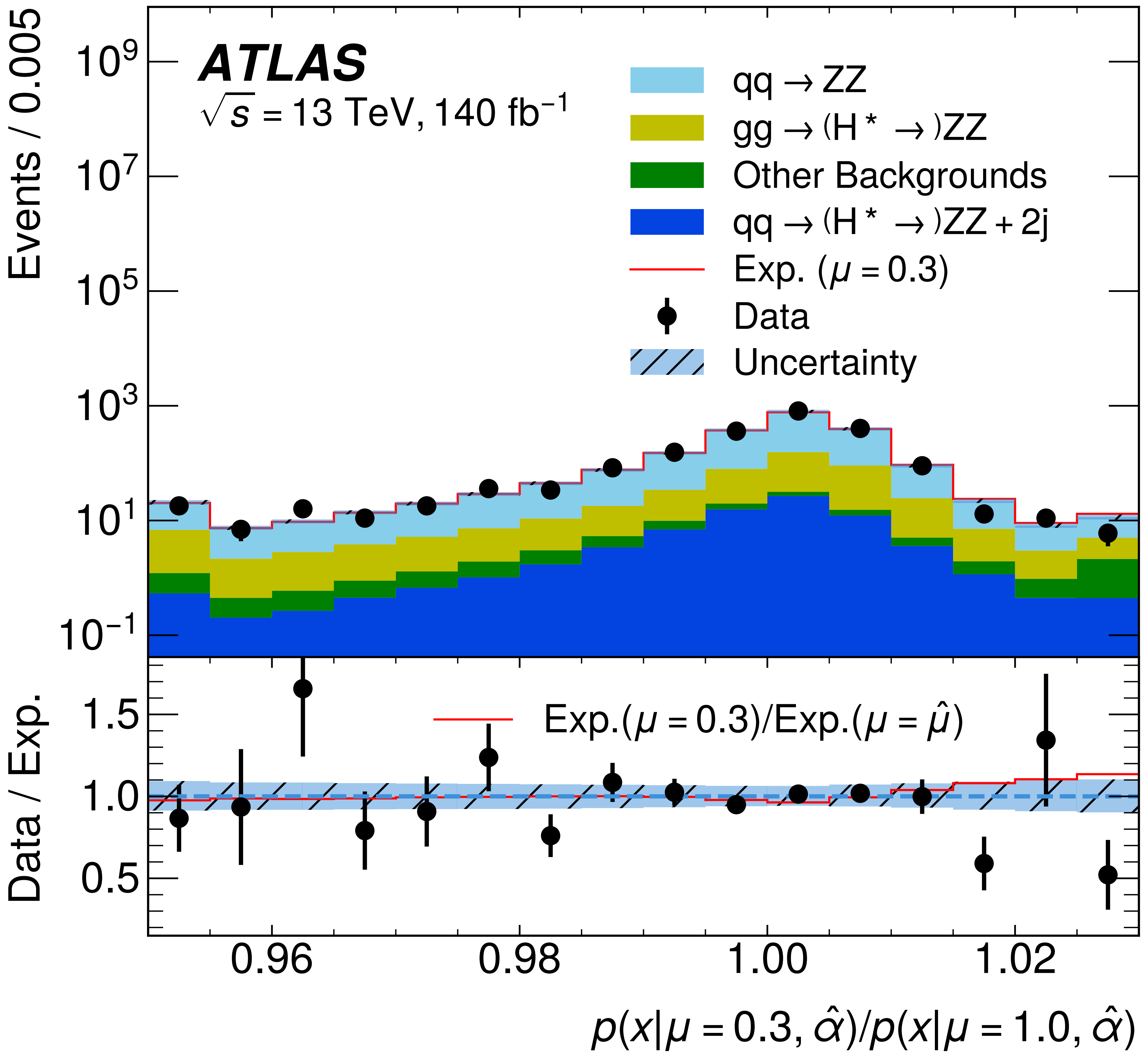


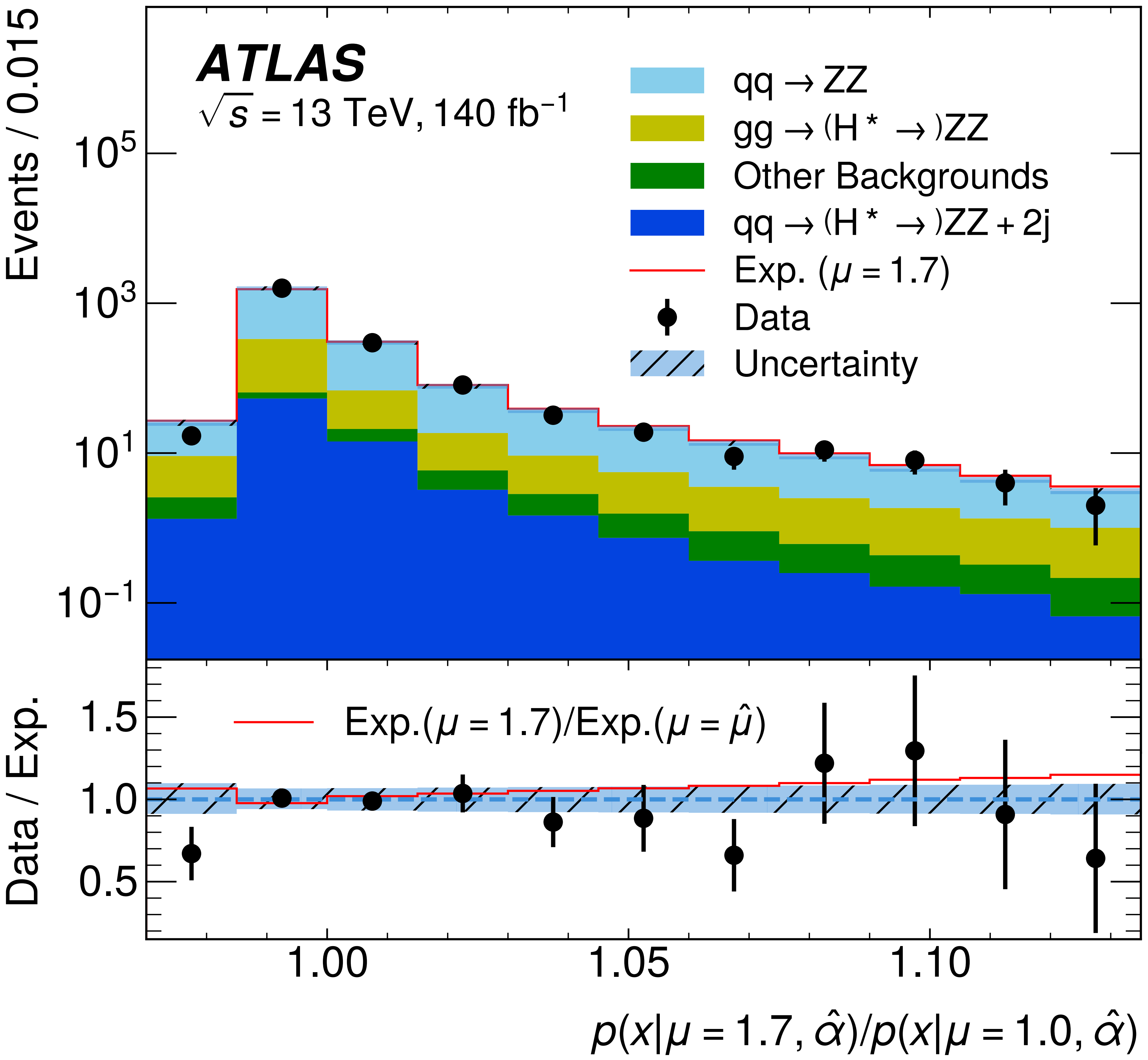
m_{4l}



Data-MC validation

Different NN observables



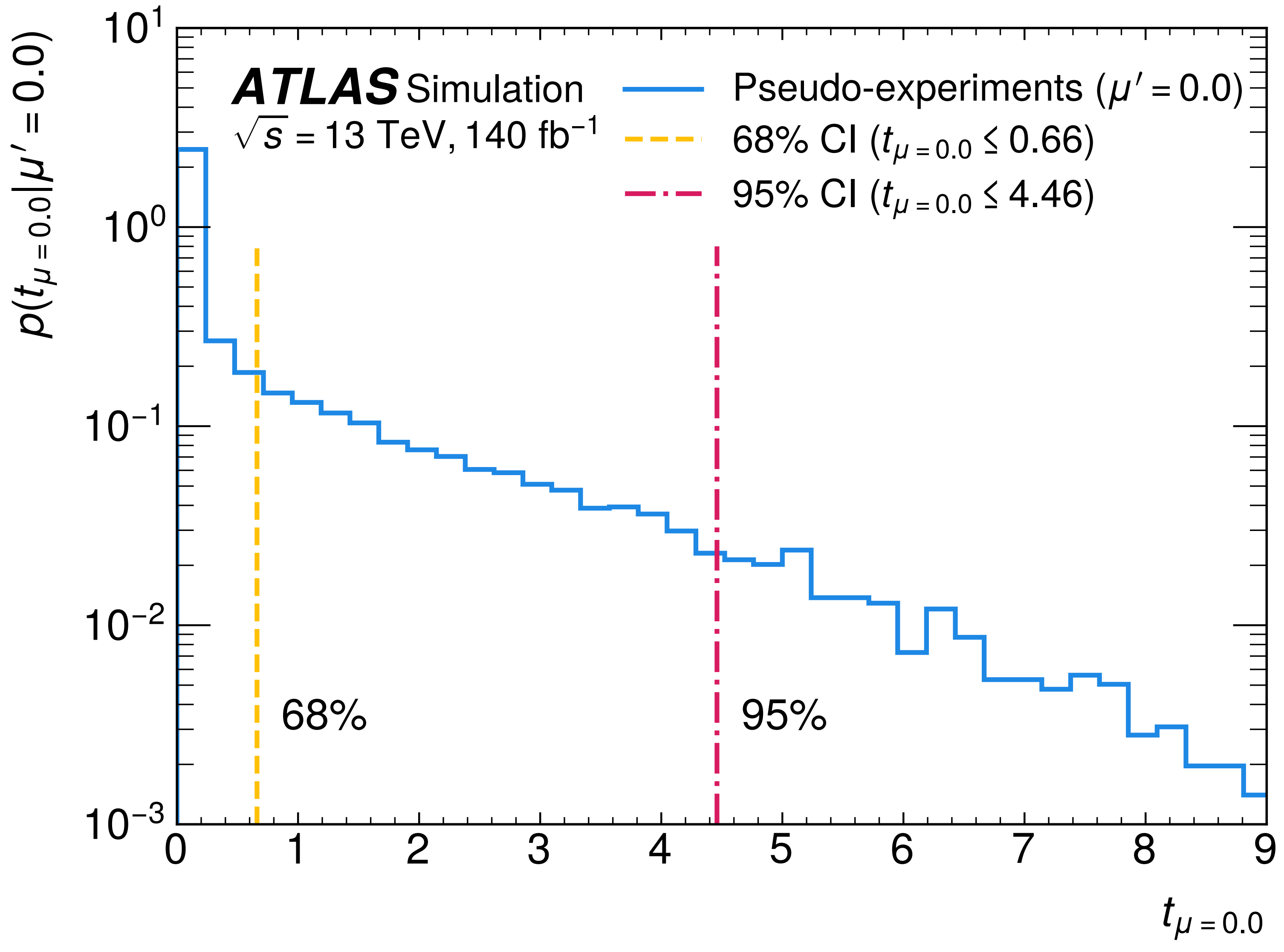


Computational Challenge: Inverting the test

- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ

Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$



t_μ

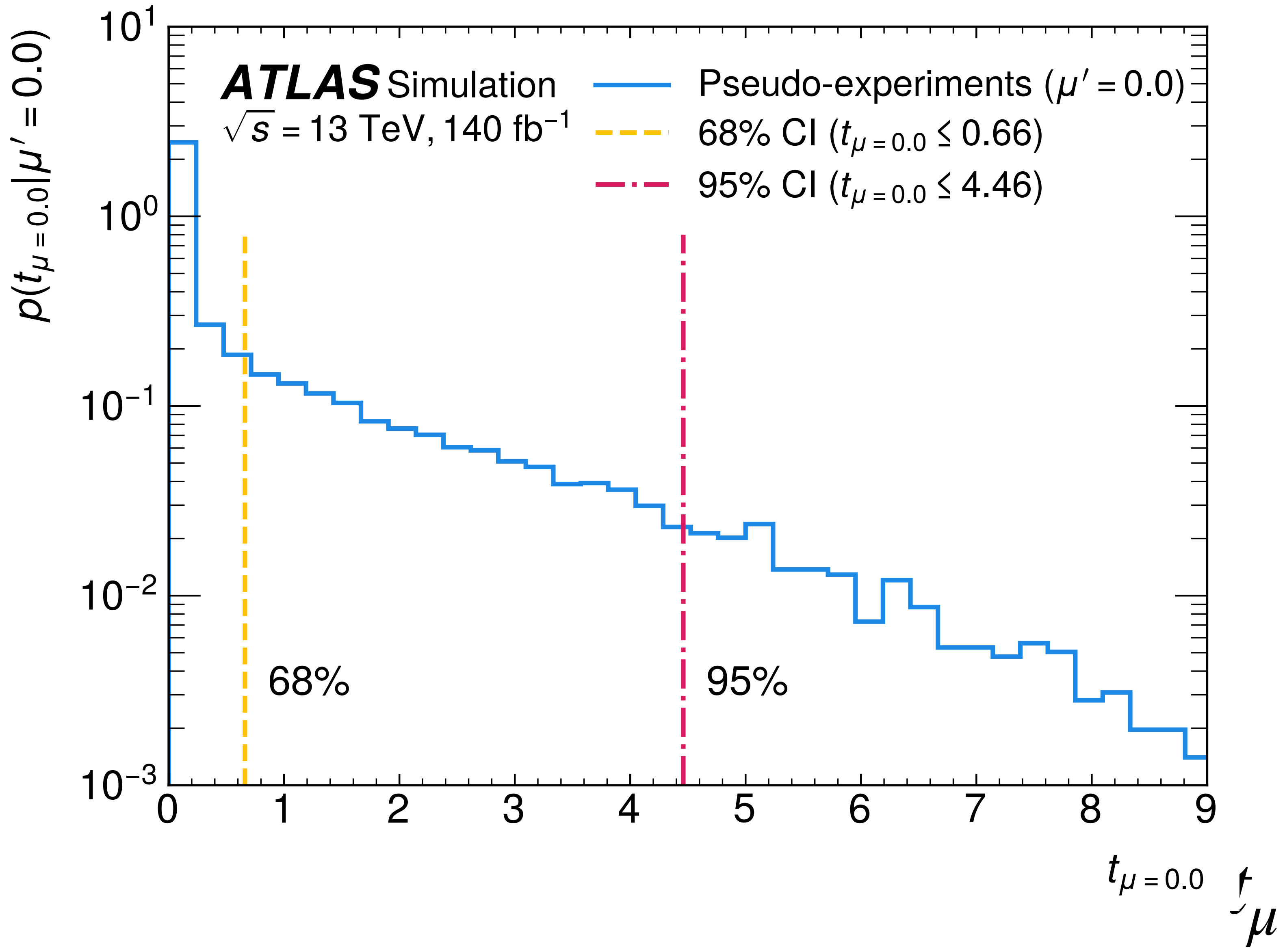
t_μ

Computational Challenge: Inverting the test

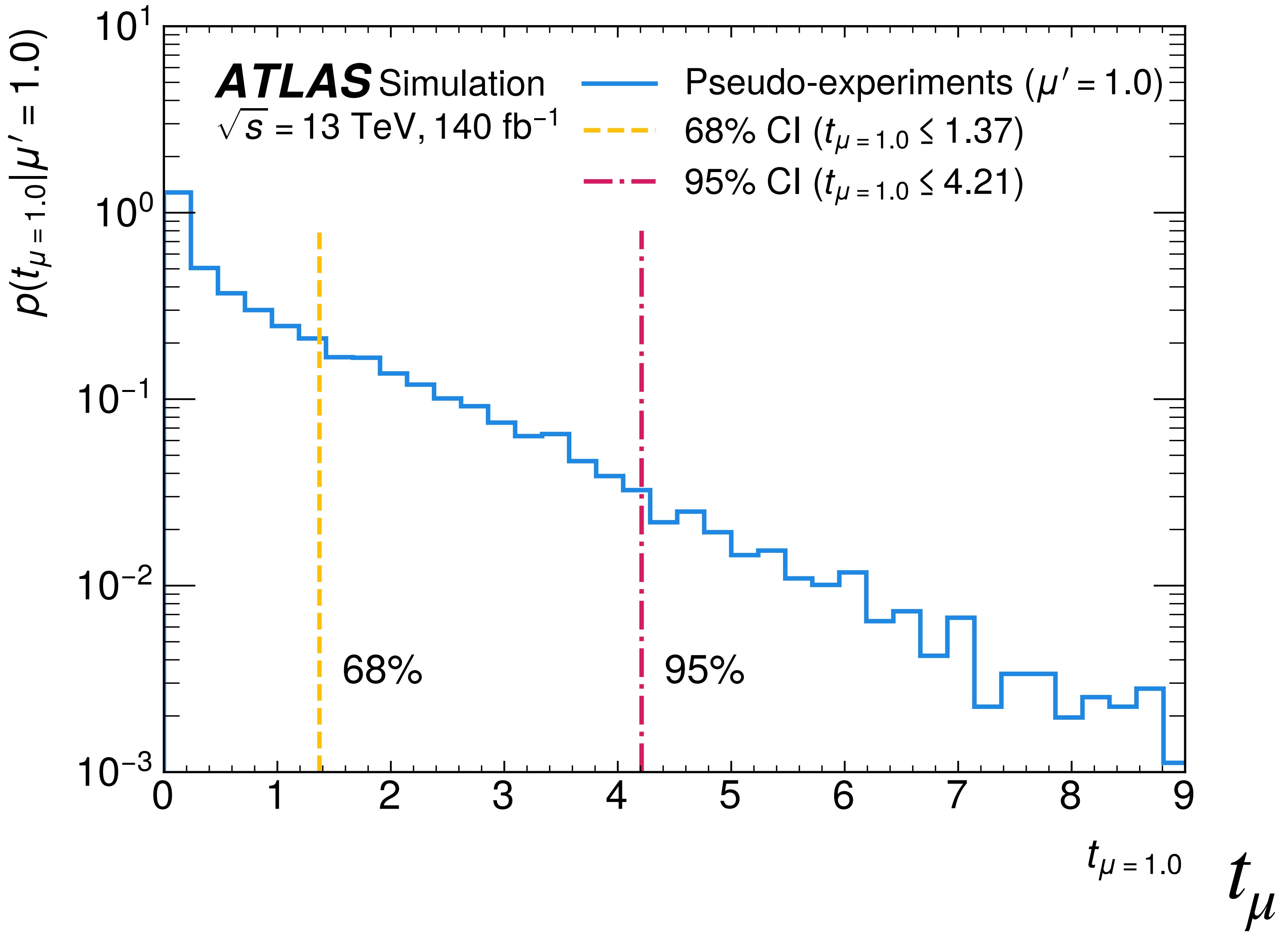
- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ

Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$

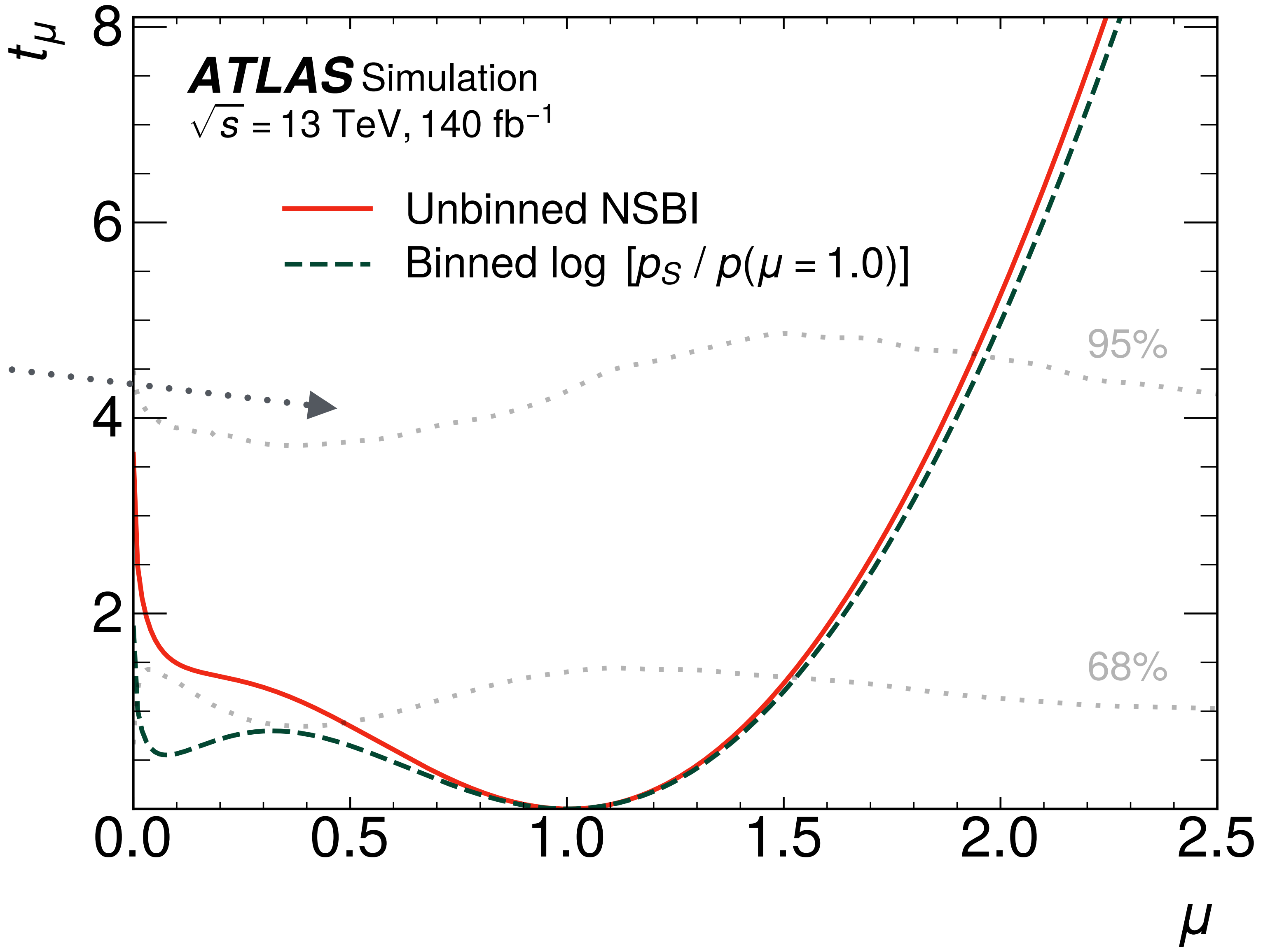


True $\mu = 1$



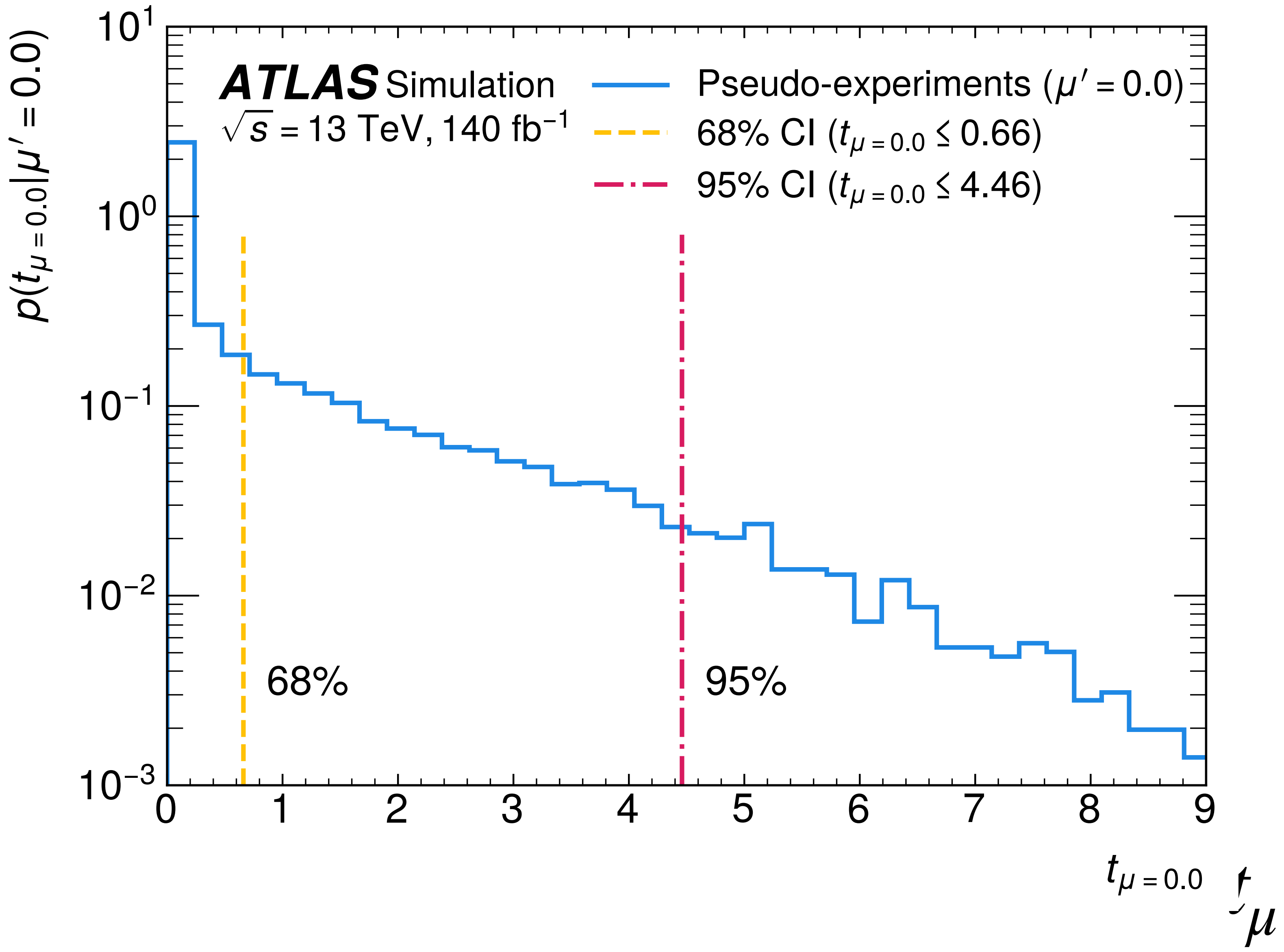
Computational Challenge: Inverting the test

- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ

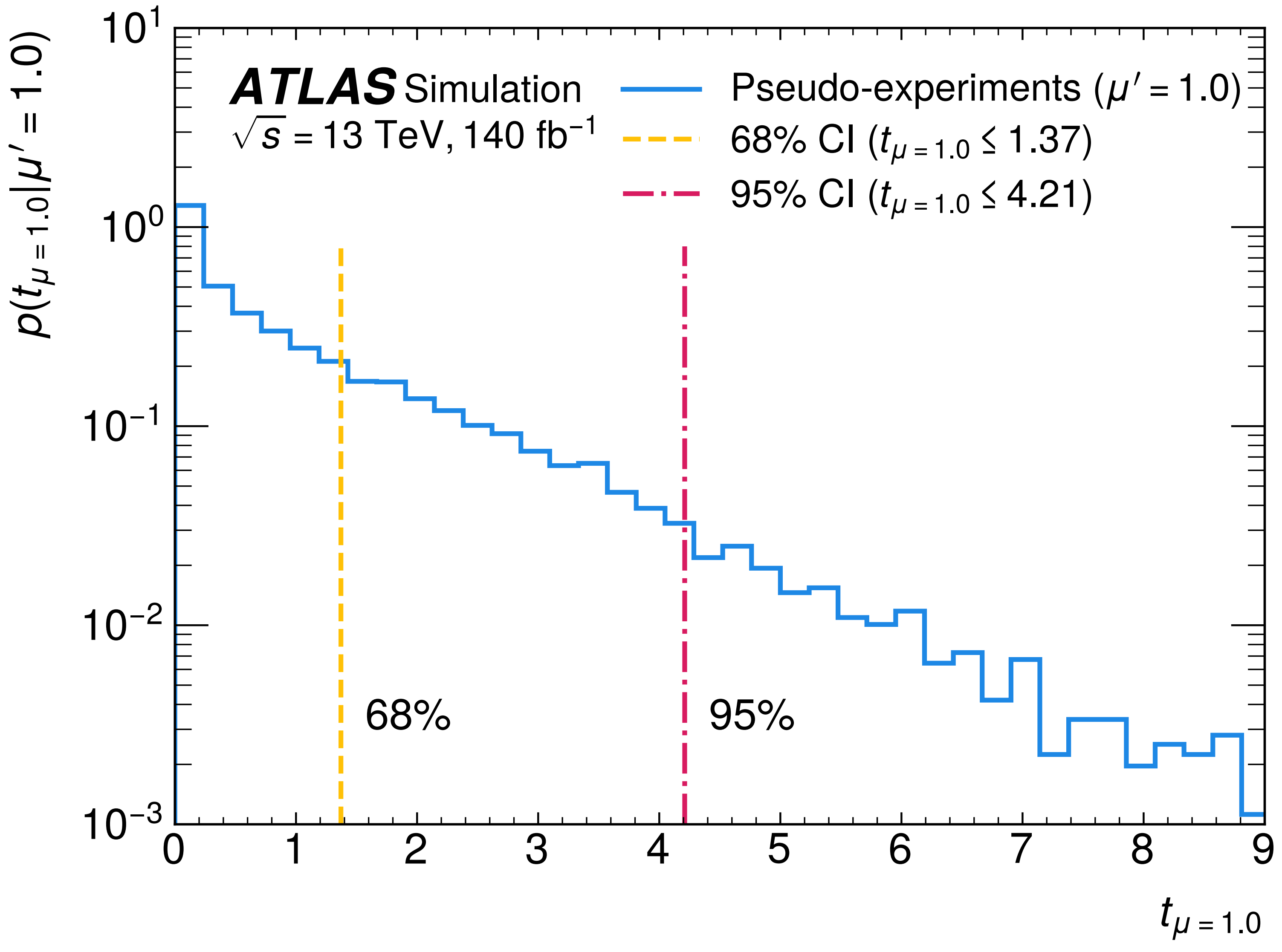


Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$



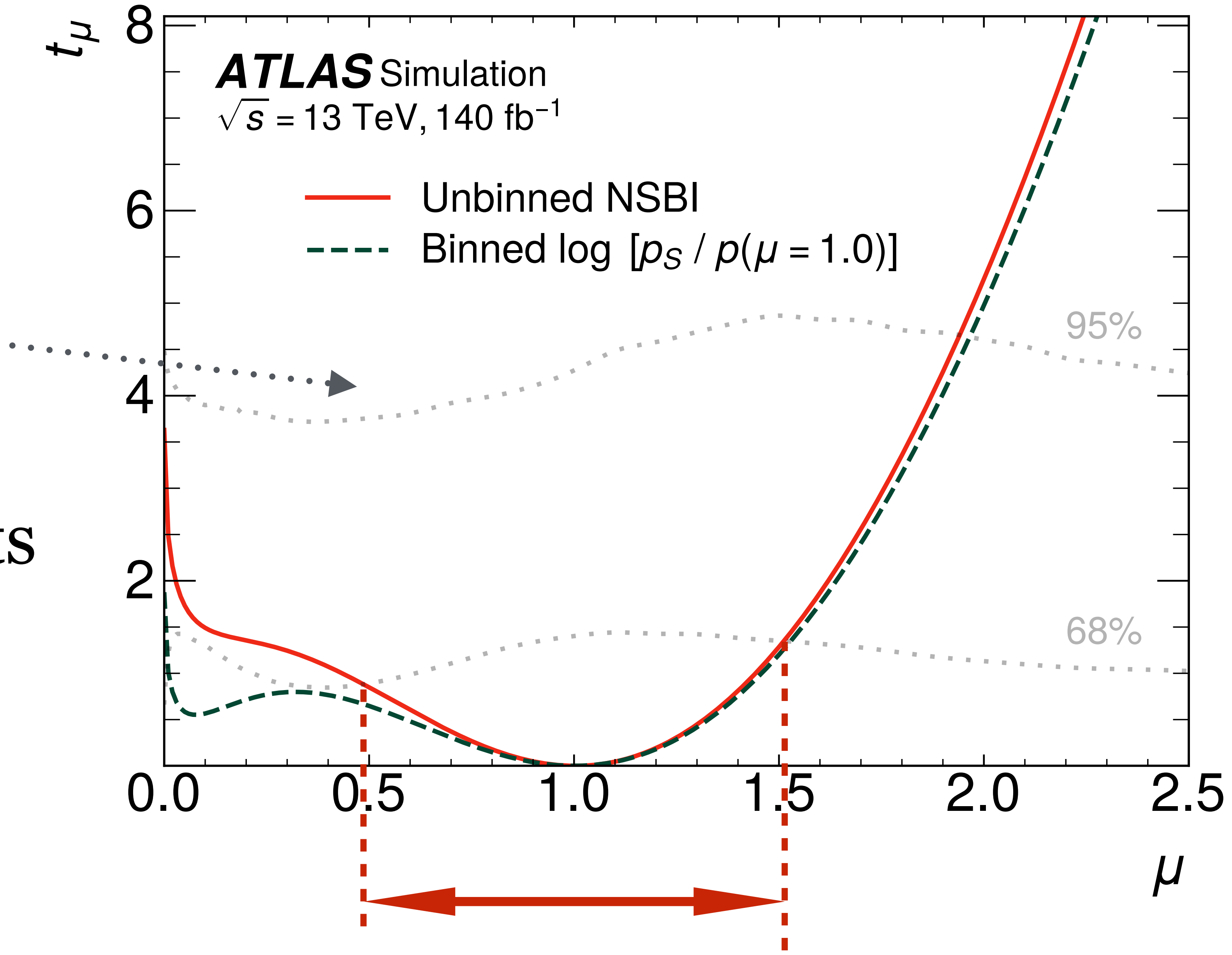
True $\mu = 1$



t_μ

Computational Challenge: Inverting the test

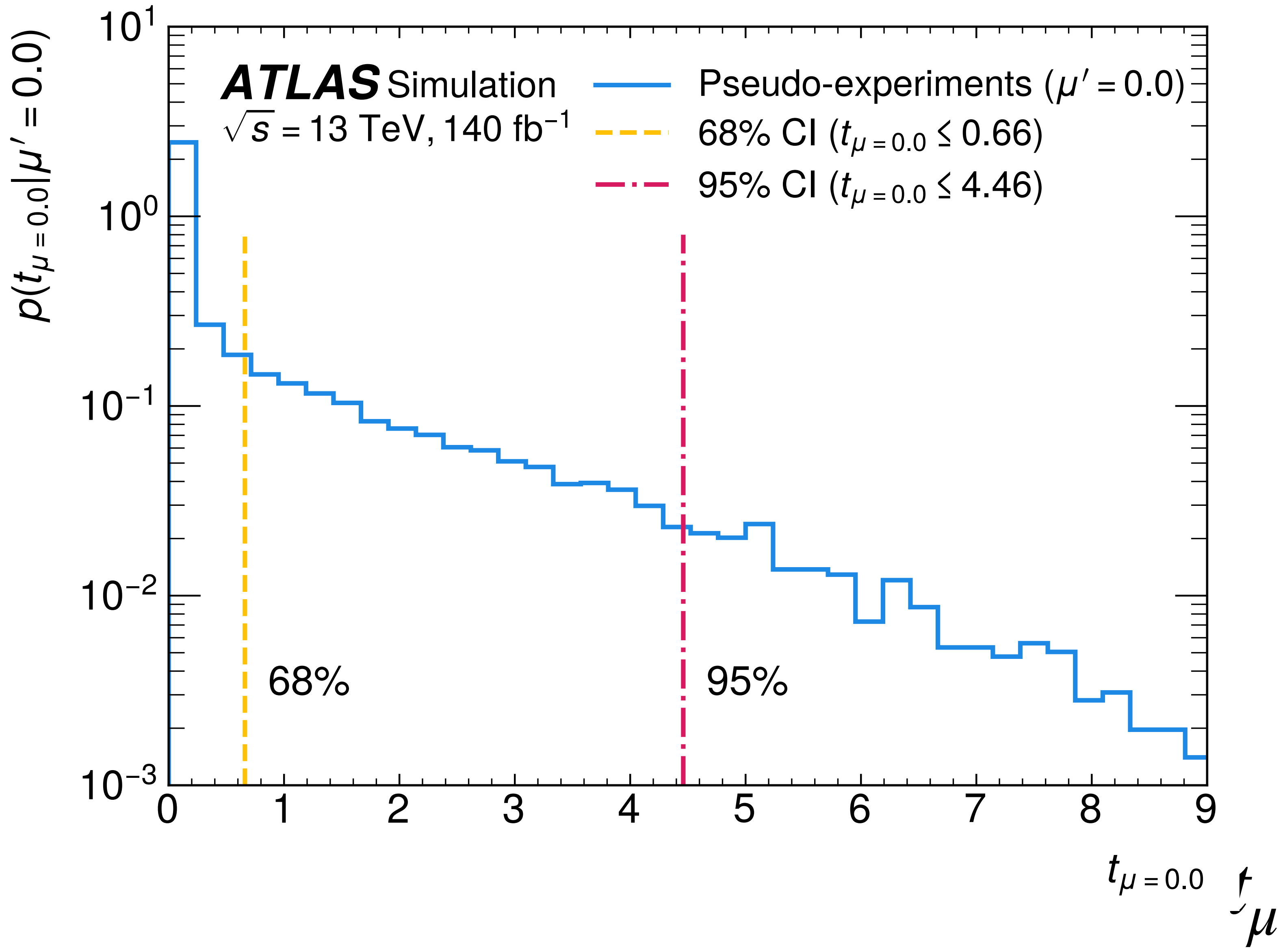
- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ



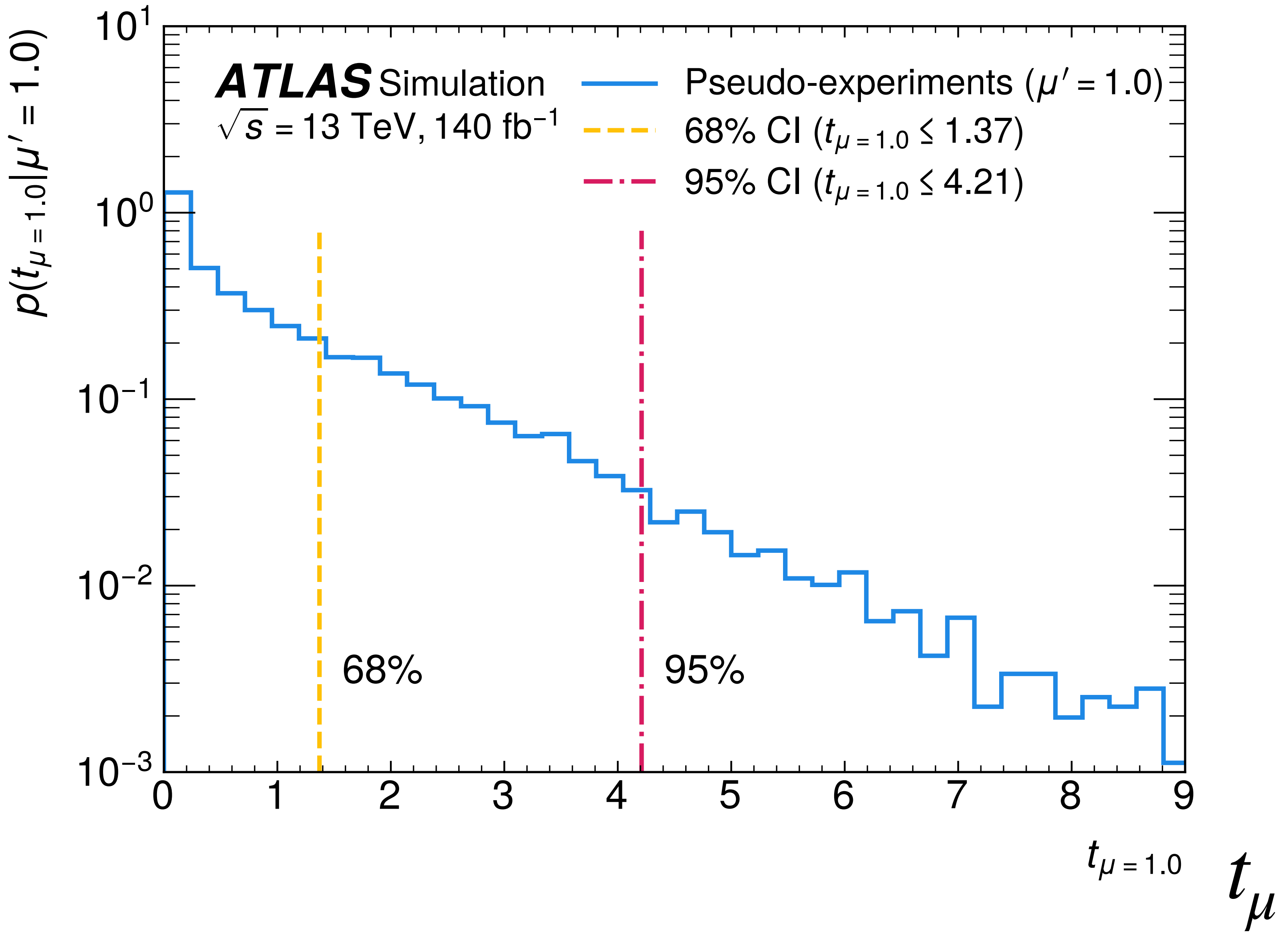
Calibrated confidence interval

Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$

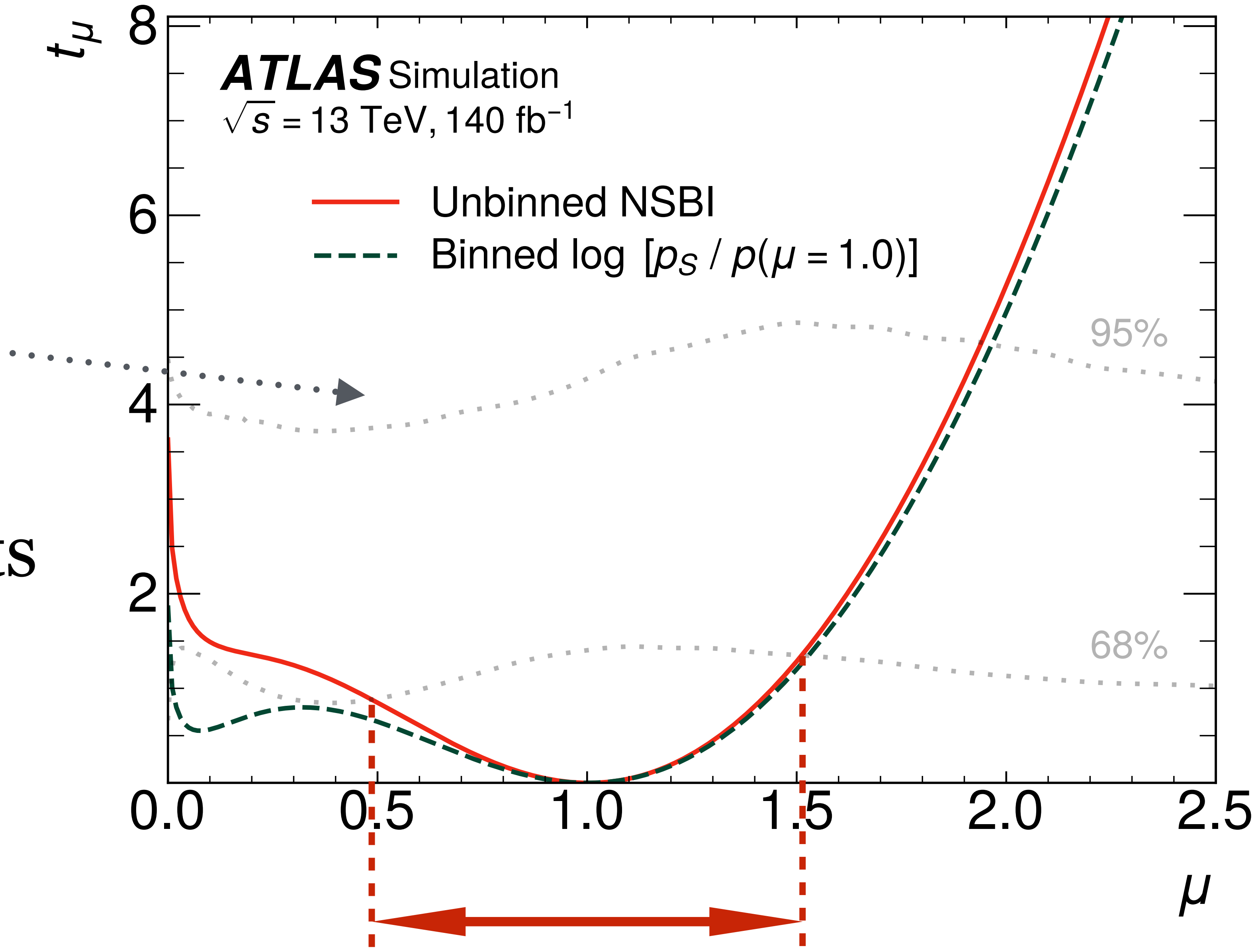


True $\mu = 1$



Computational Challenge: Inverting the test

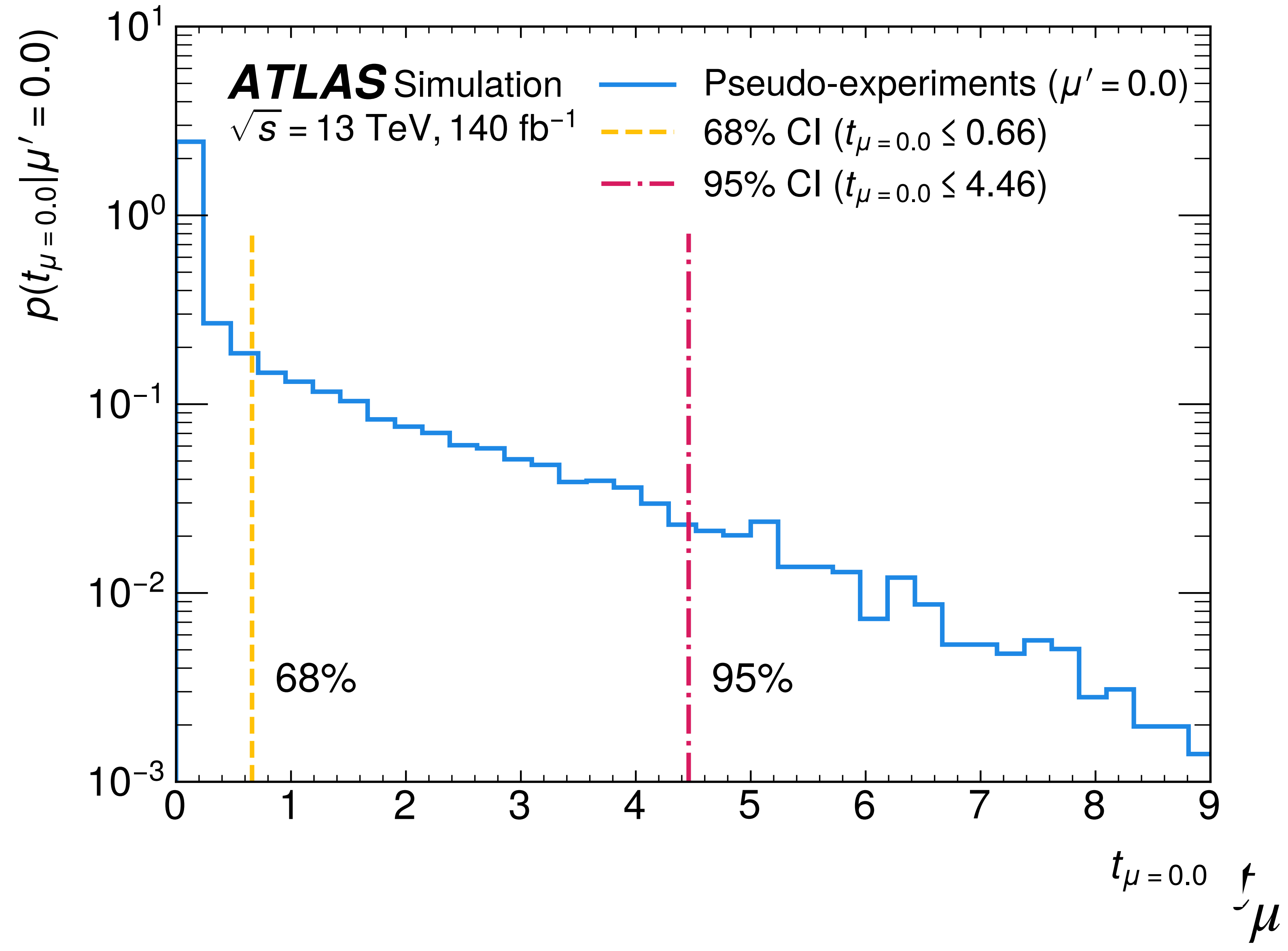
- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ



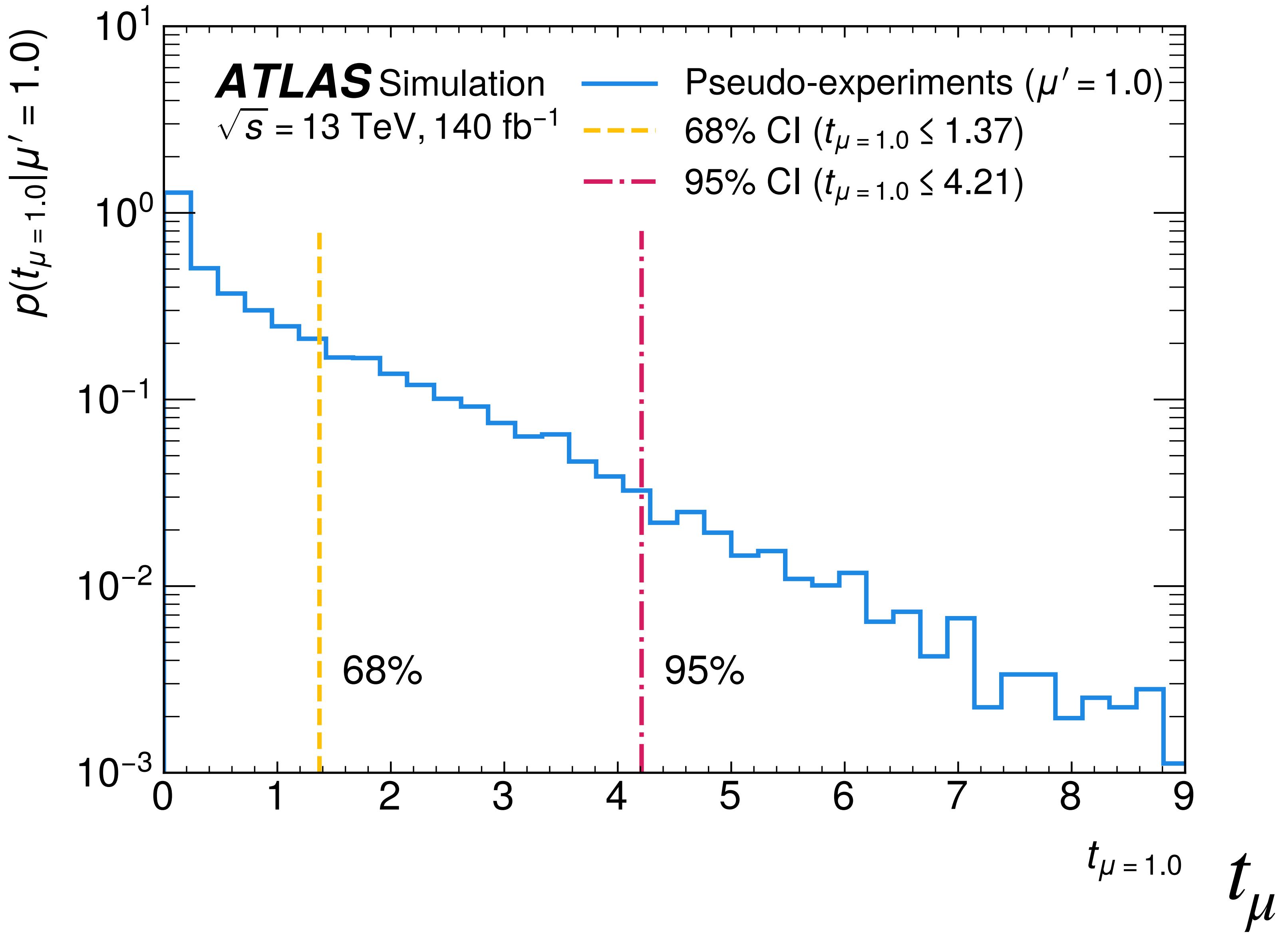
Calibrated confidence interval

Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

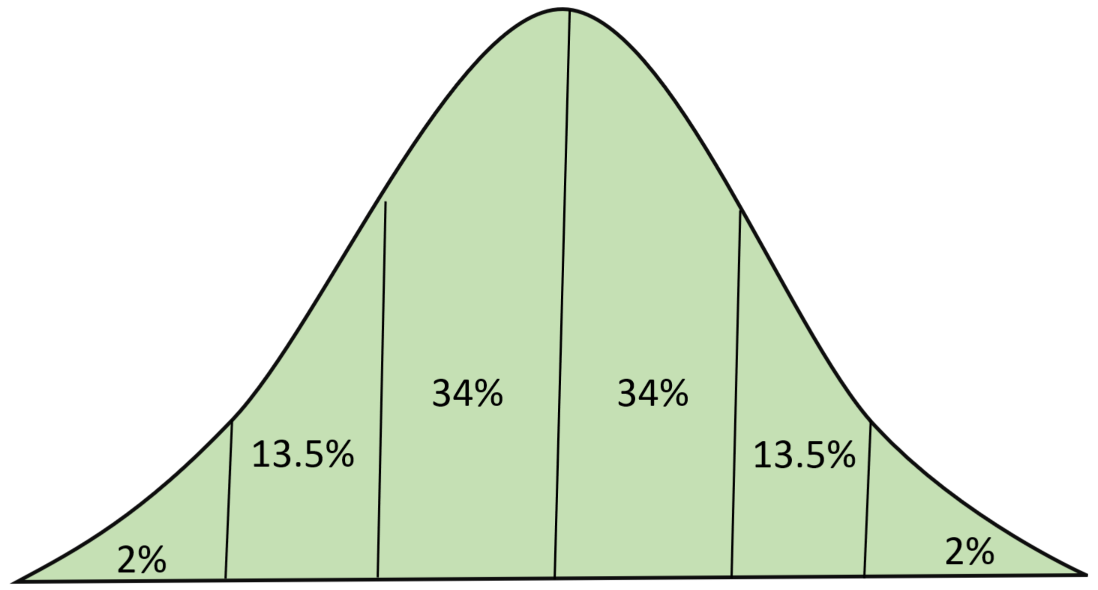
True $\mu = 0$



True $\mu = 1$



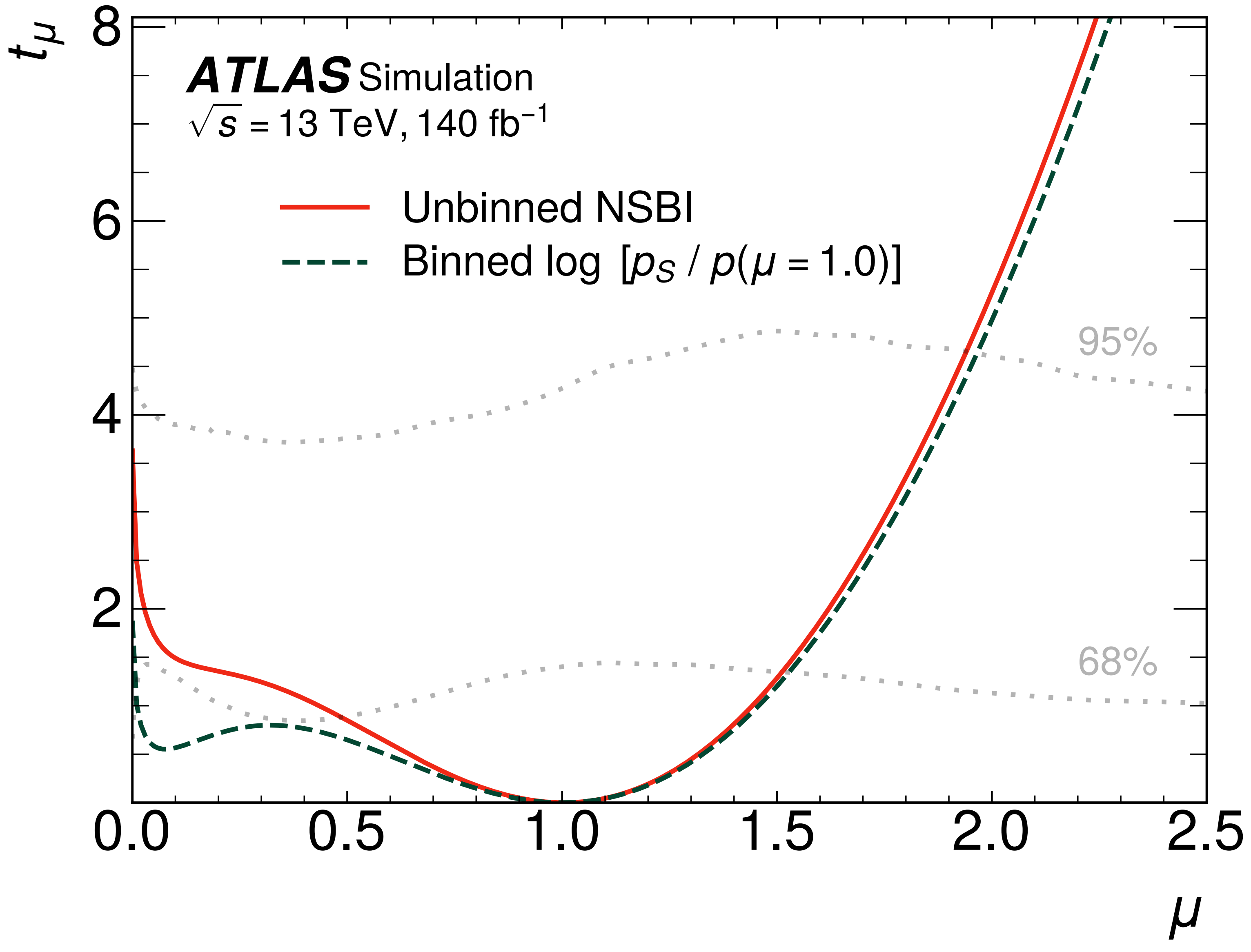
Task is to determine quantiles



Fast confidence belts with QR

Neyman construction guarantees accuracy of confidence intervals, even if you don't trust the NN estimates

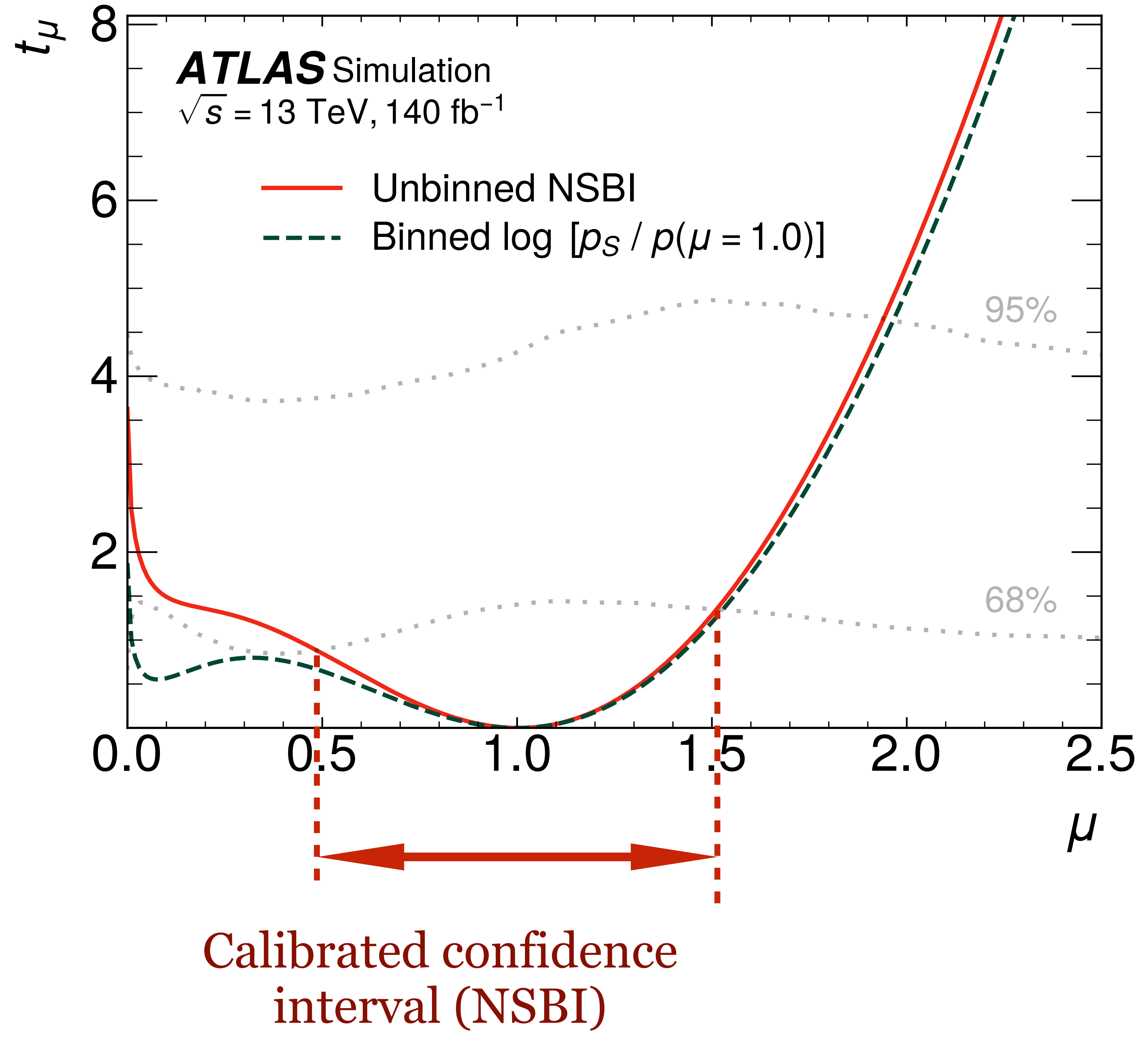
But very slow and expensive



Fast confidence belts with QR

Neyman construction guarantees accuracy of confidence intervals, even if you don't trust the NN estimates

But very slow and expensive

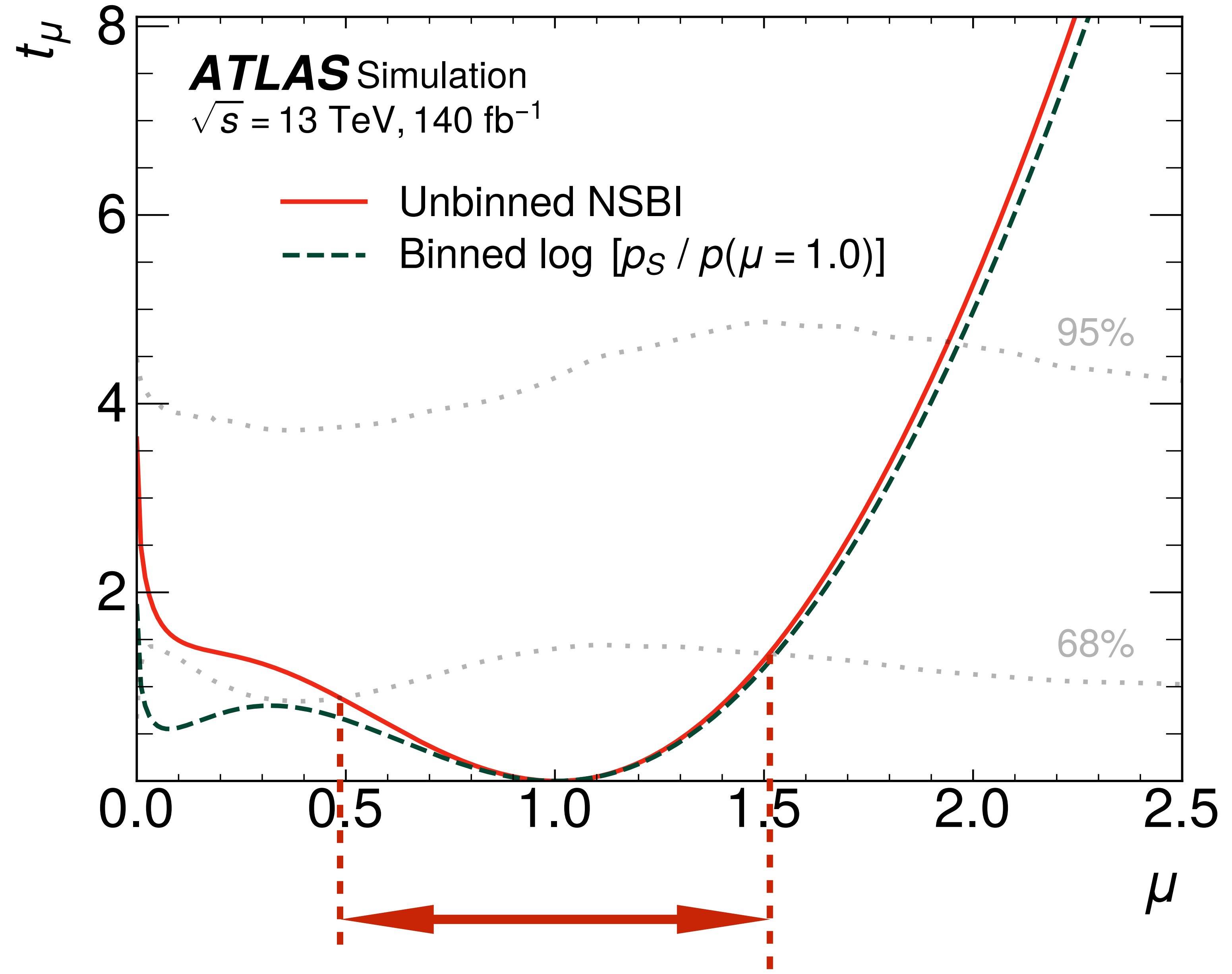


Fast confidence belts with QR

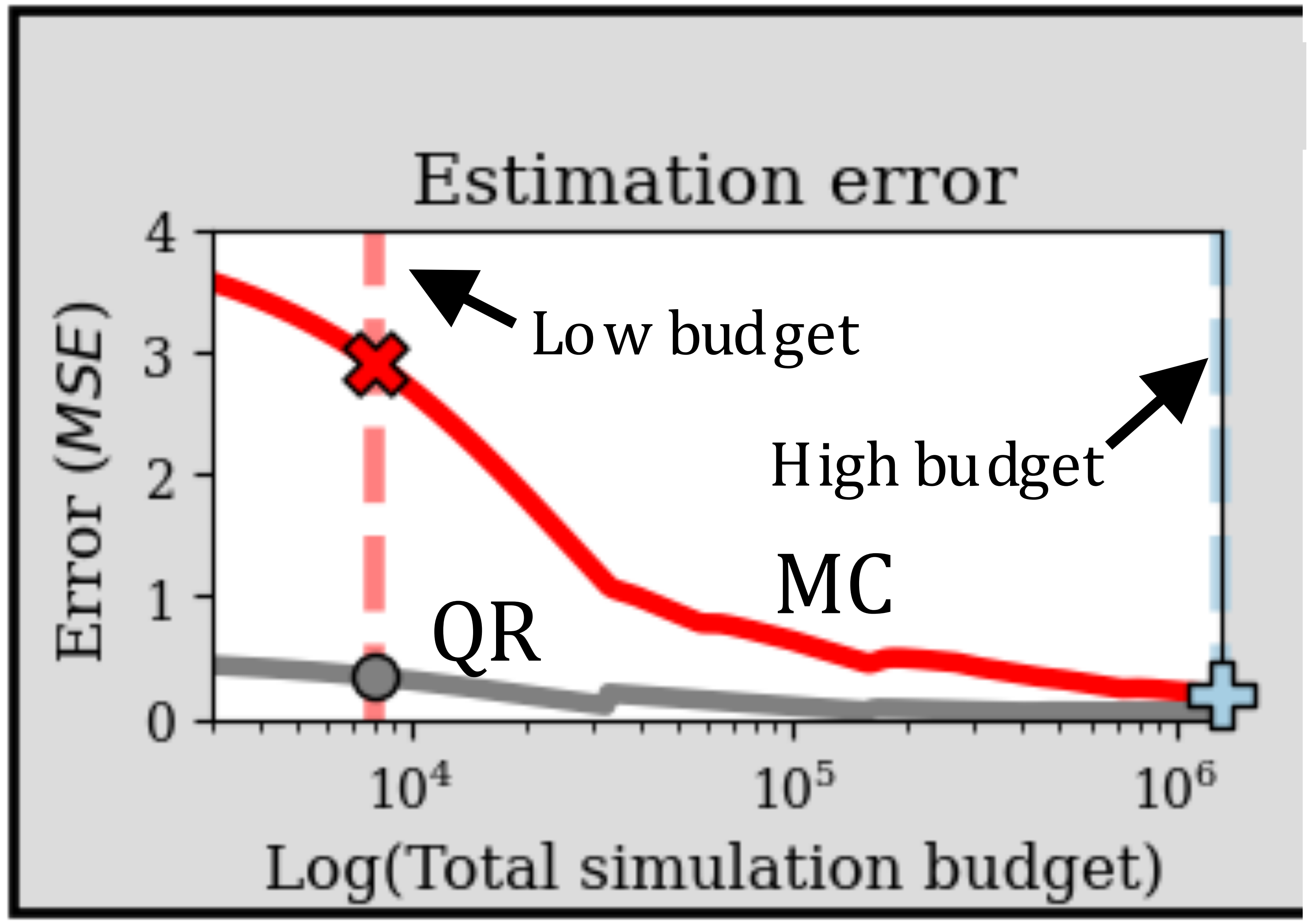
Neyman construction guarantees accuracy of confidence intervals, even if you don't trust the NN estimates

But very slow and expensive

Can do it faster and more accurately with quantile regression (QR)



Calibrated confidence interval (NSBI)

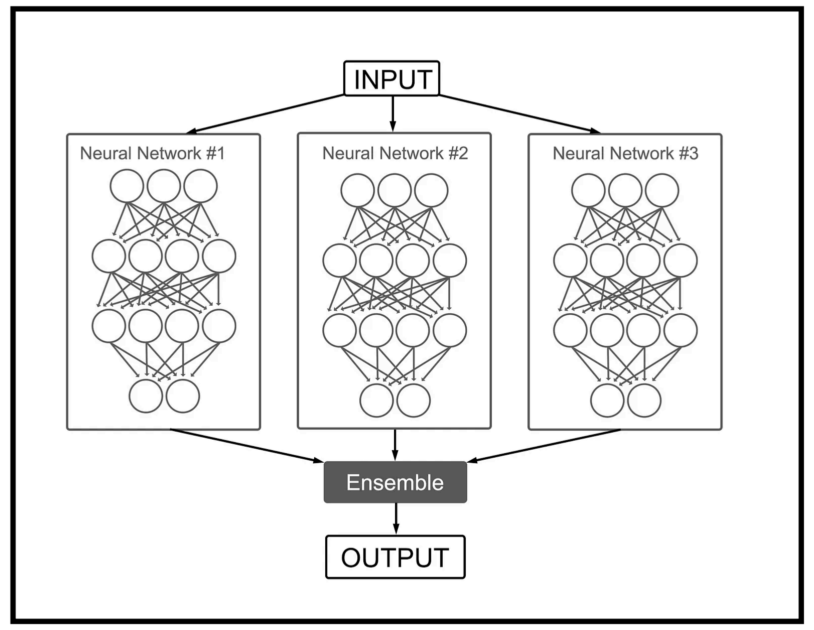


Workflows for the computational challenge

- HEP performs ~100 measurements per year that could benefit
- Careful design and validation of 10,000 networks per measurement is infeasible

Need fully automated workflows with Ray:

Train ensembles on HPC



Details: HEP-CCE, Bollweg et al (incl. Ghosh)

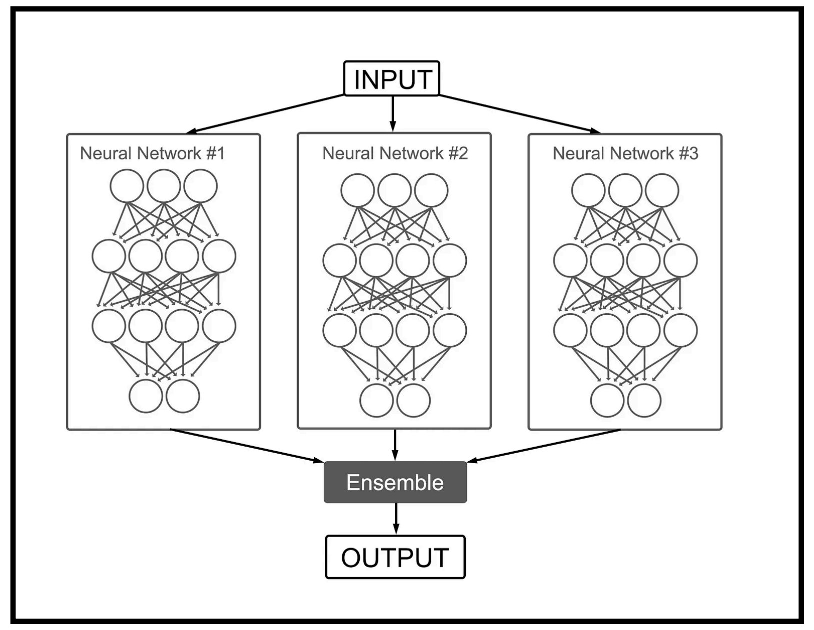
IRIS-HEP Tool (Sandesara et al): Unbinned NSBI fits

Workflows for the computational challenge

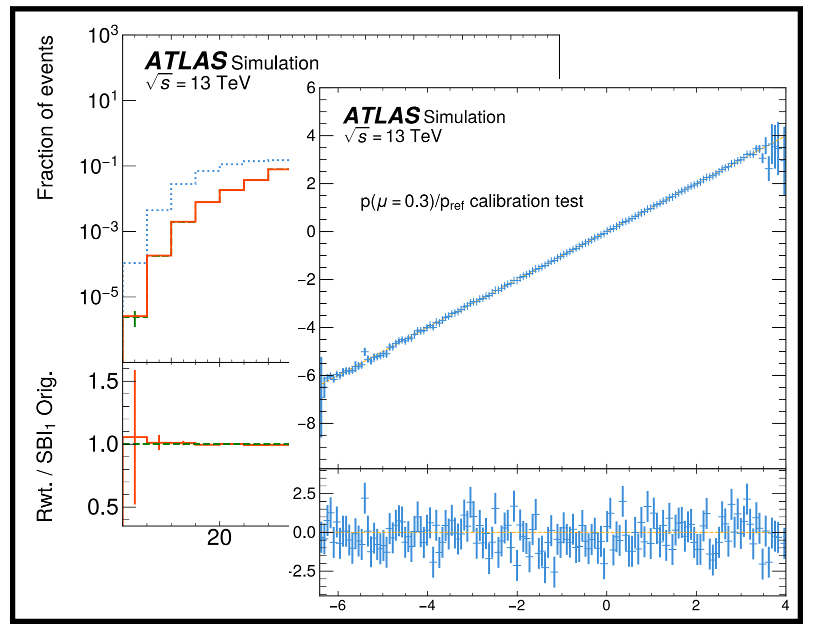
- HEP performs ~100 measurements per year that could benefit
- Careful design and validation of 10,000 networks per measurement is infeasible

Need fully automated workflows with Ray:

Train ensembles on HPC



Validation of density estimates



Details: HEP-CCE, Bollweg et al (incl. Ghosh)

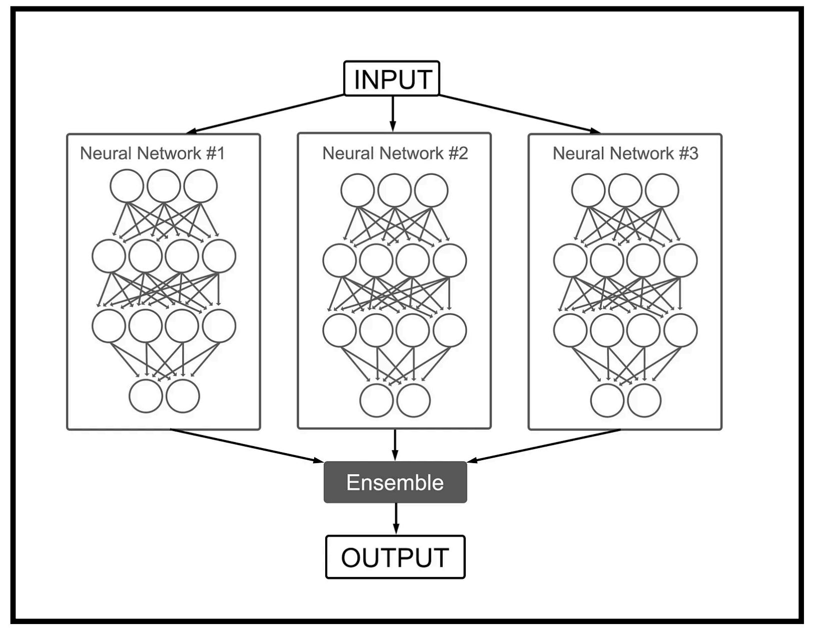
IRIS-HEP Tool (Sandesara et al): Unbinned NSBI fits

Workflows for the computational challenge

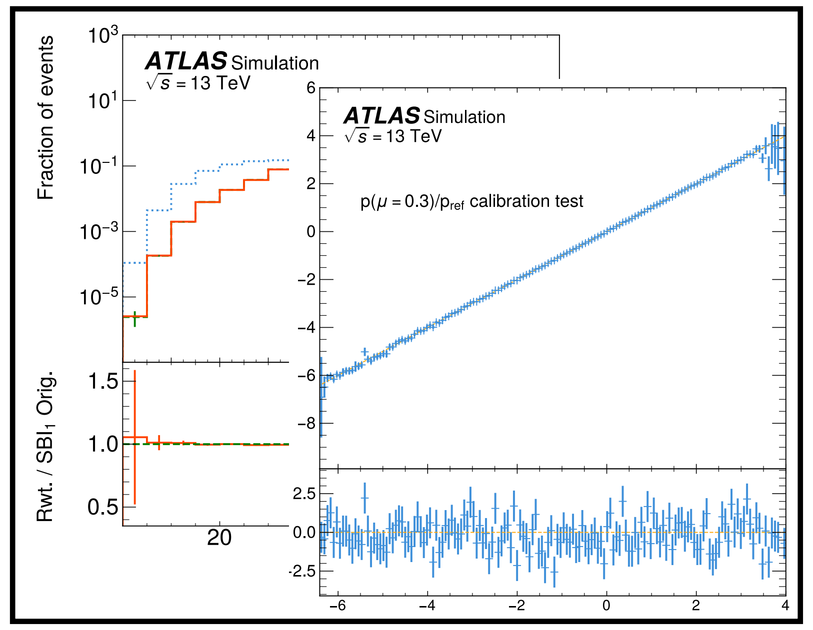
- HEP performs ~100 measurements per year that could benefit
- Careful design and validation of 10,000 networks per measurement is infeasible

Need fully automated workflows with Ray:

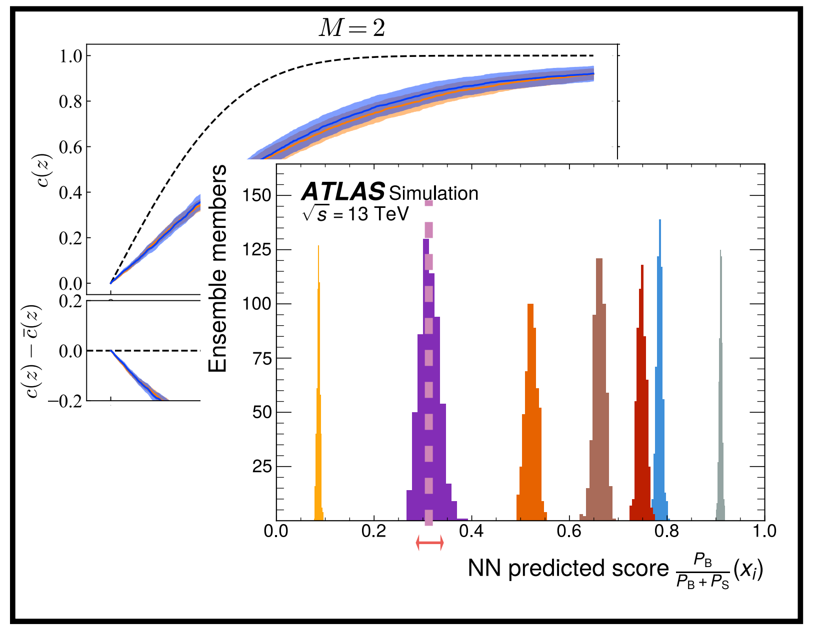
Train ensembles on HPC



Validation of density estimates



Uncertainty quantification



Details: HEP-CCE, Bollweg et al (incl. Ghosh)

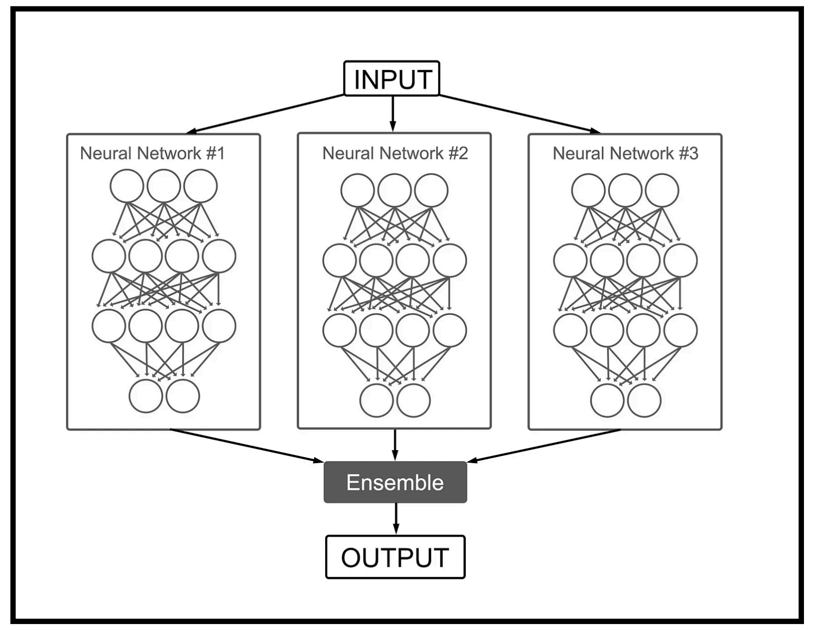
IRIS-HEP Tool (Sandesara et al): Unbinned NSBI fits

Workflows for the computational challenge

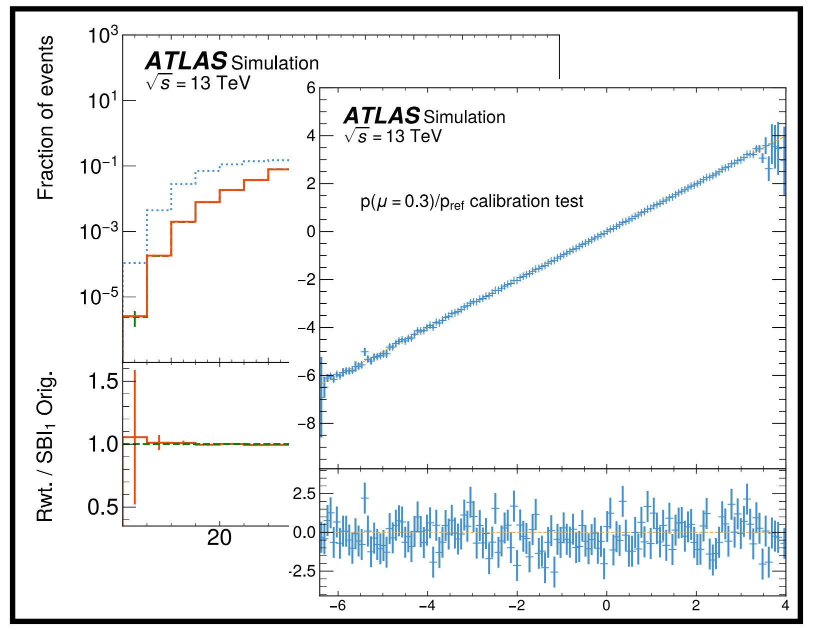
- HEP performs ~100 measurements per year that could benefit
- Careful design and validation of 10,000 networks per measurement is infeasible

Need fully automated workflows with Ray:

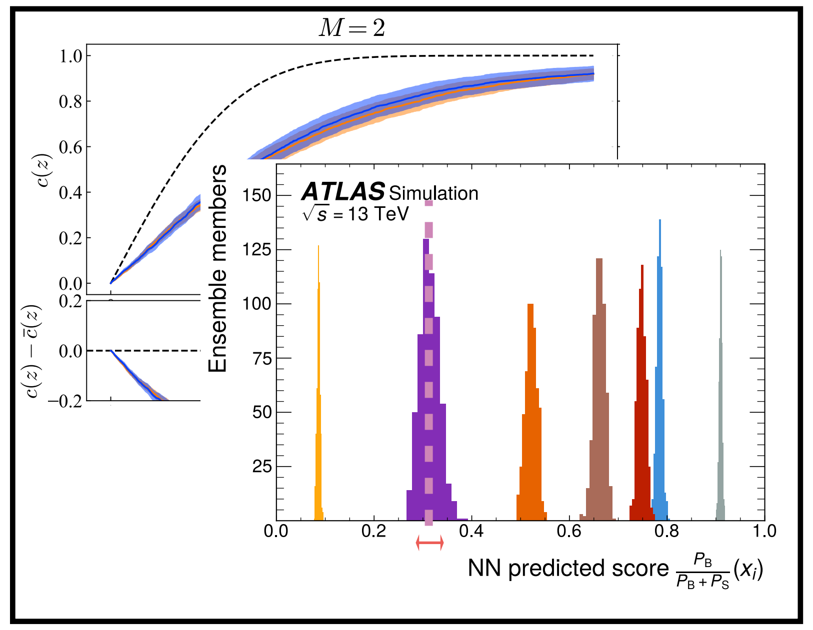
Train ensembles on HPC



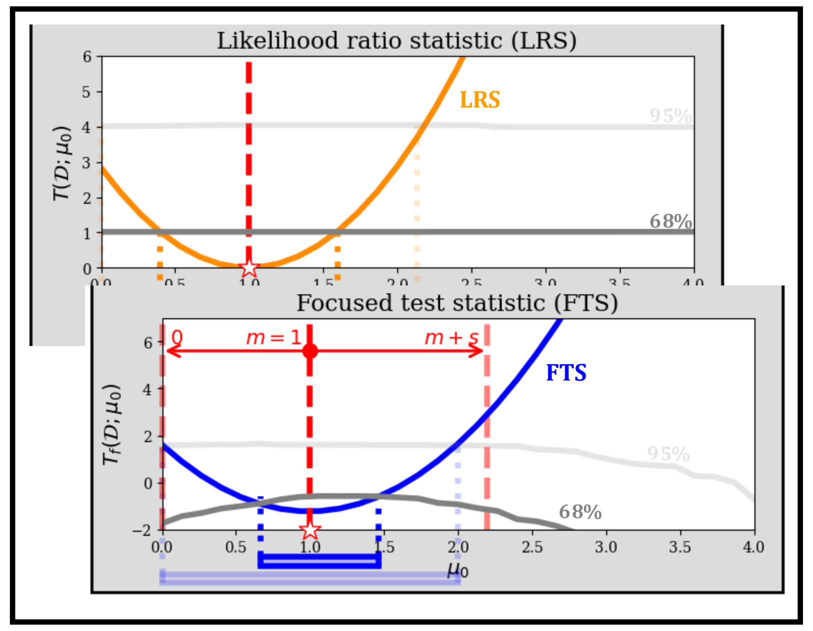
Validation of density estimates



Uncertainty quantification



Neyman inversion with pseudo-experiments



Details: HEP-CCE, Bollweg et al (incl. Ghosh)

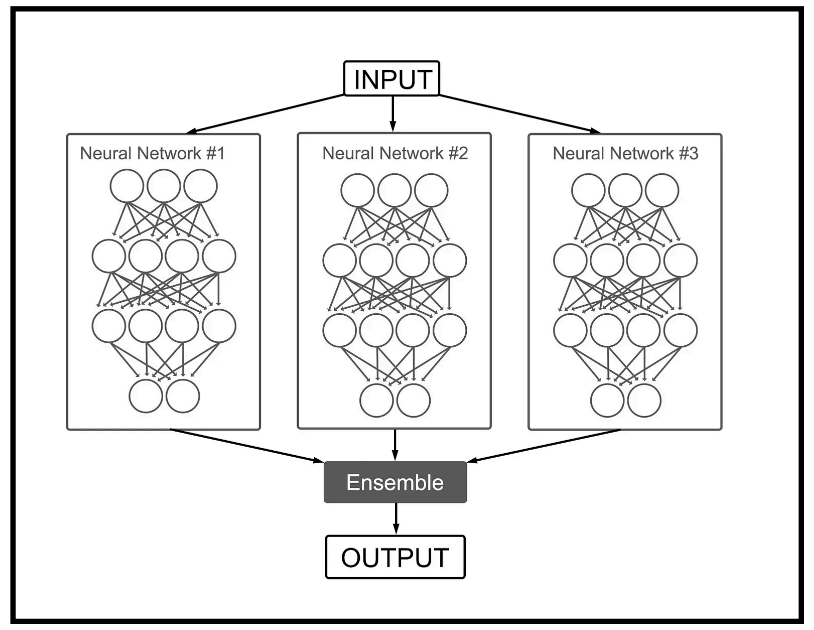
IRIS-HEP Tool (Sandesara et al): Unbinned NSBI fits

Workflows for the computational challenge

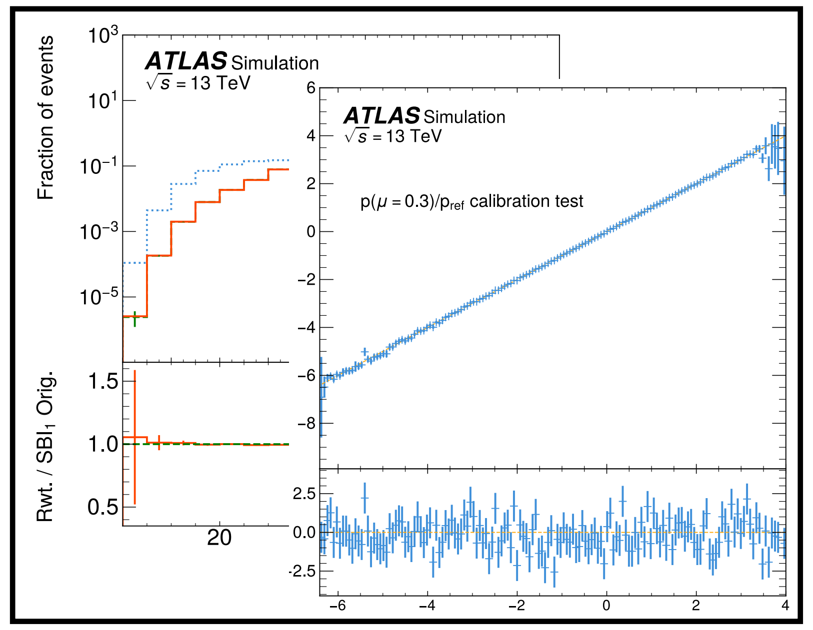
- HEP performs ~100 measurements per year that could benefit
- Careful design and validation of 10,000 networks per measurement is infeasible

Need fully automated workflows with Ray:

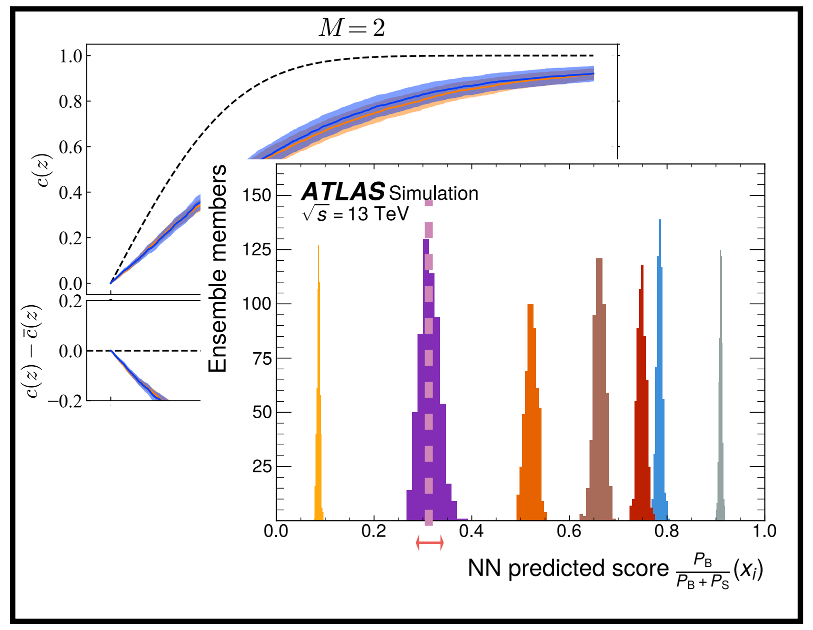
Train ensembles on HPC



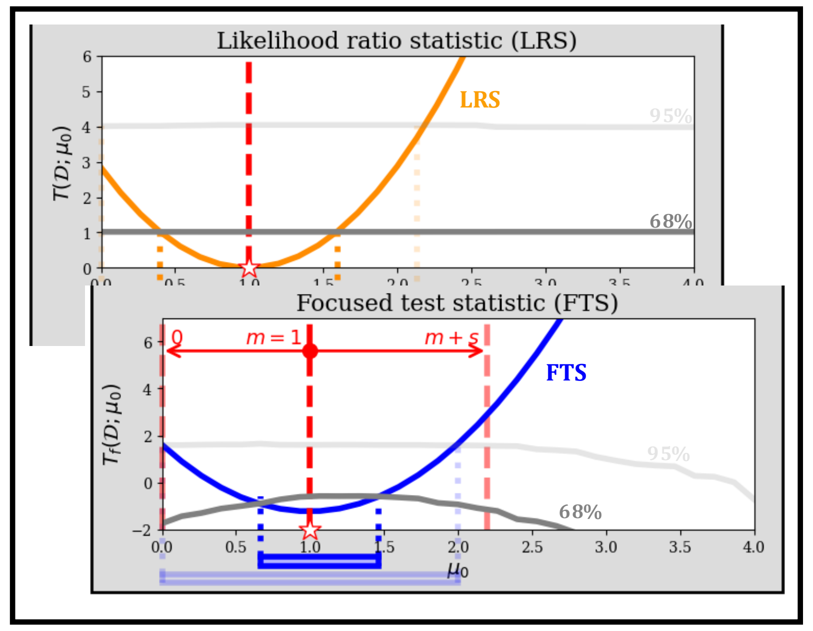
Validation of density estimates



Uncertainty quantification



Neyman inversion with pseudo-experiments



Optimize

Details: HEP-CCE, Bollweg et al (incl. Ghosh)

IRIS-HEP Tool (Sandesara et al): Unbinned NSBI fits

Until now, we have replaced individual pieces with ML in age-old likelihood ratio test

Do we dare question the test itself?

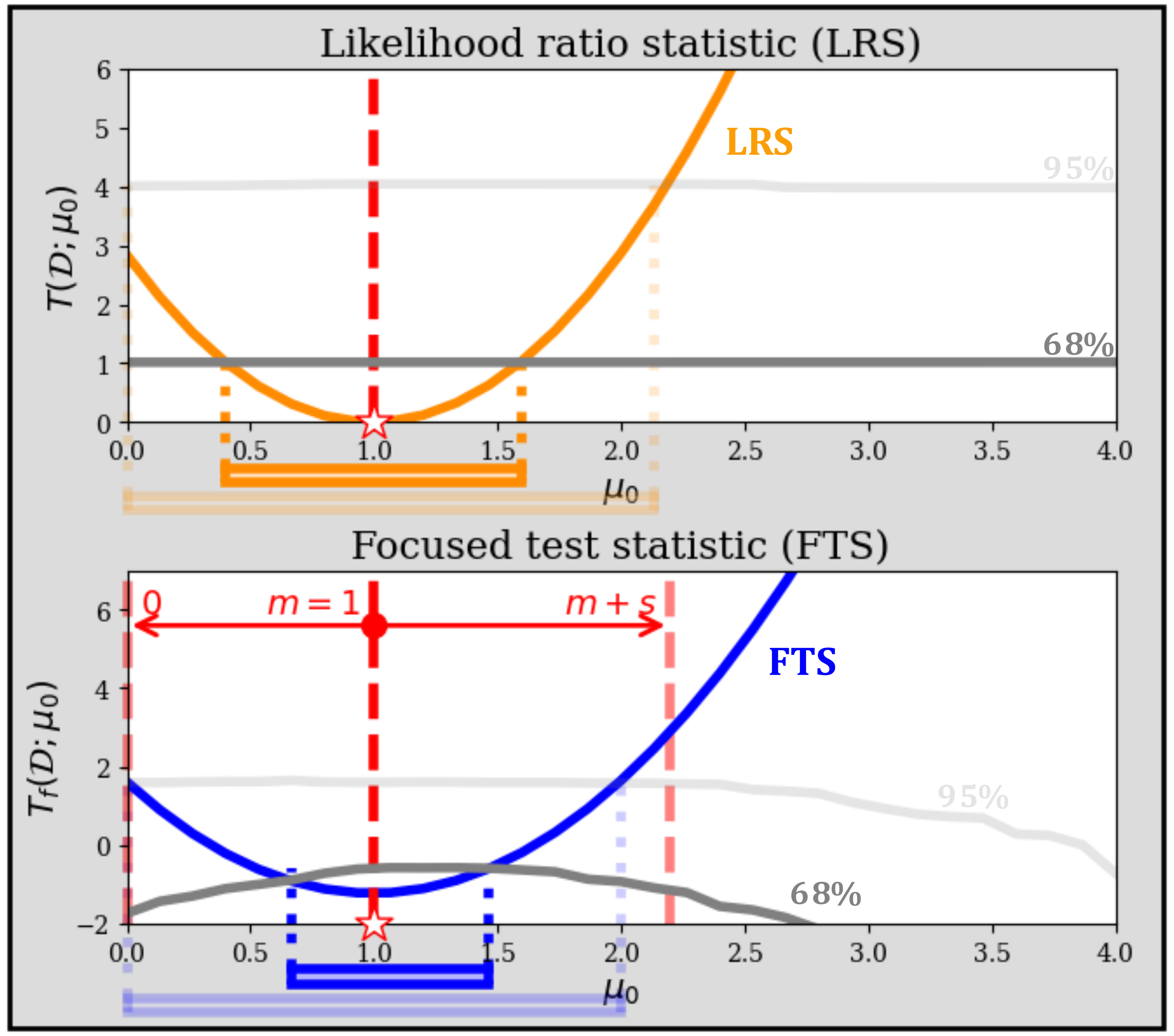
Until now, we have replaced individual pieces with ML in age-old likelihood ratio test

Do we dare question the test itself?

You *can* beat the likelihood ratio test statistic

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \longrightarrow FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

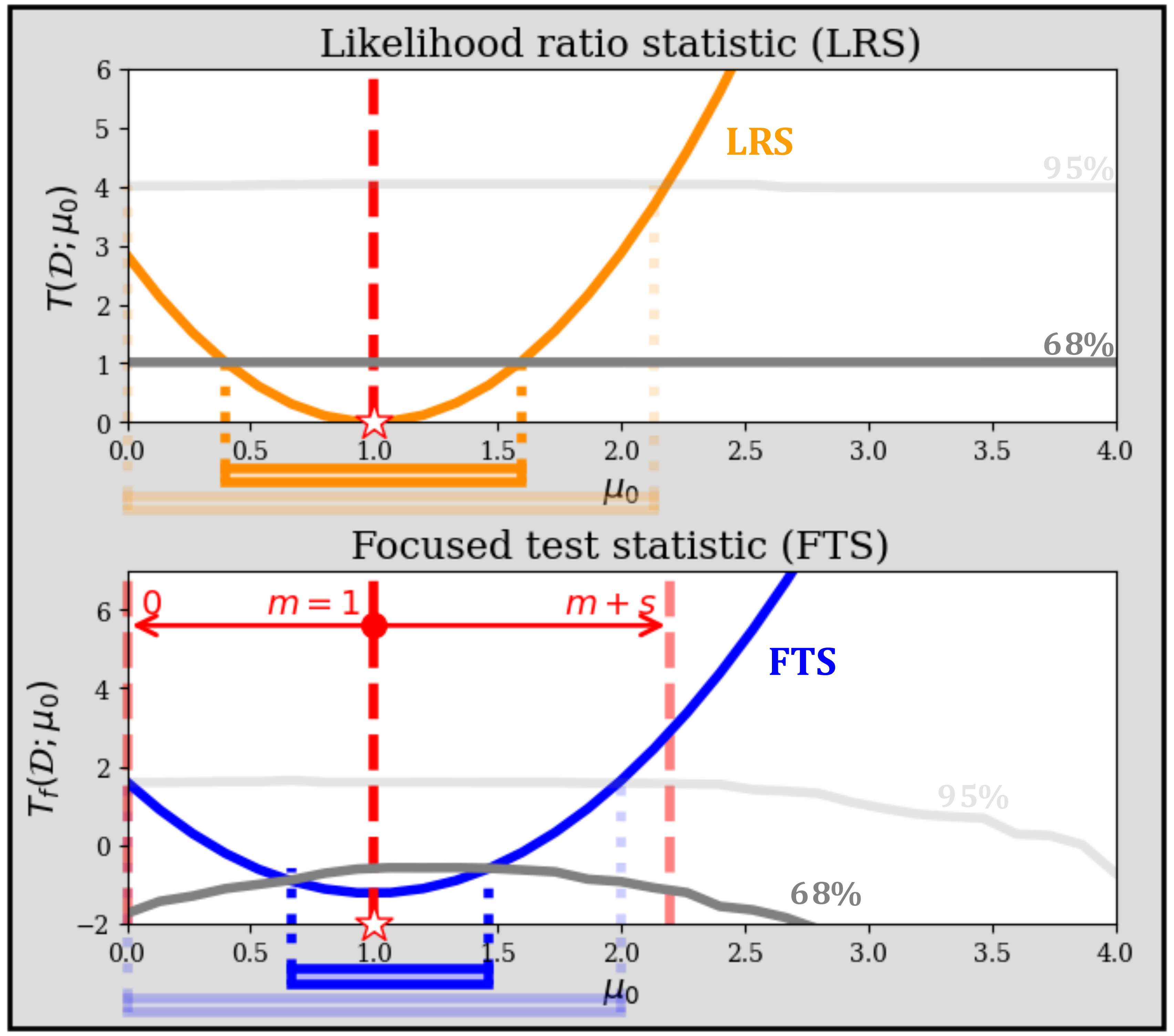
- Denominator in 'Focused Test Statistic' (FTS) knows about all alternate hypotheses



You *can* beat the likelihood ratio test statistic

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \longrightarrow FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

- Denominator in 'Focused Test Statistic' (FTS) knows about all alternate hypotheses

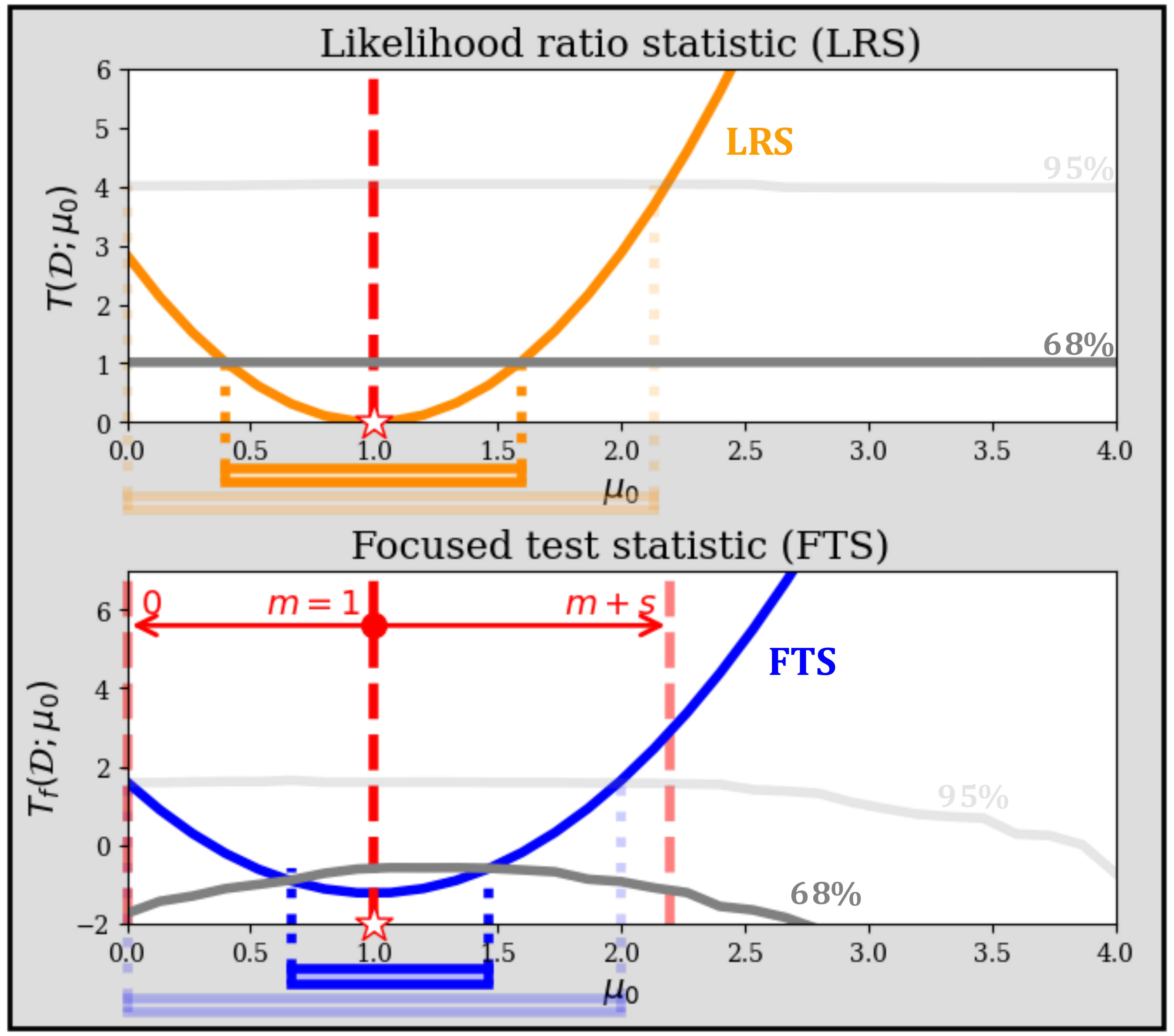


Shorter median length for confidence intervals with FTS even in 'asymptotic regime' where Wilks' applies

You *can* beat the likelihood ratio test statistic

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \longrightarrow FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

- Denominator in ‘Focused Test Statistic’ (FTS) knows about all alternate hypotheses
- $f(\mu)$ focuses statistical power in meaningful regions of parameter space
  - Particularly useful in small sample / small signal regime
- Fast critical value estimation with ML

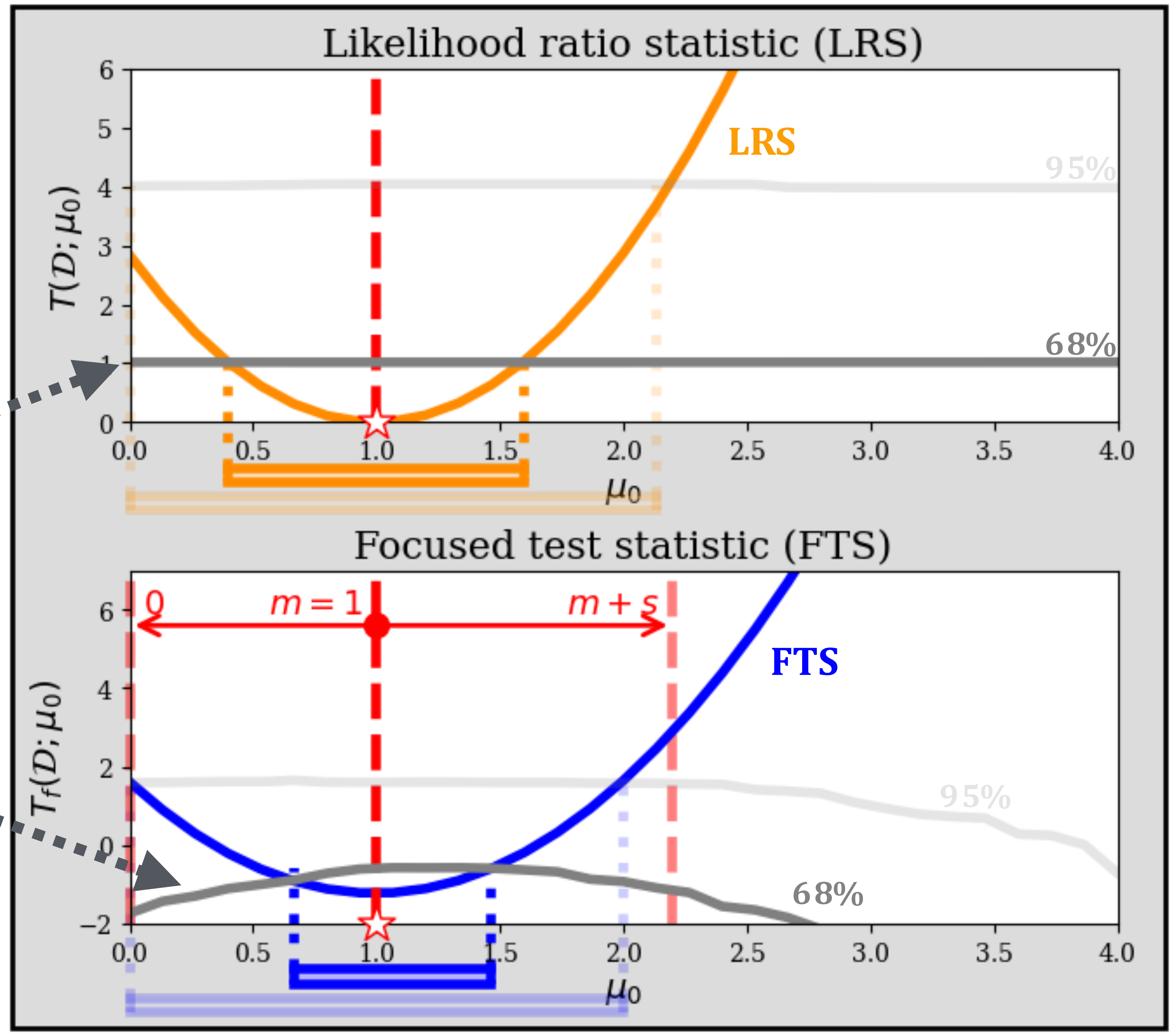


Shorter median length for confidence intervals with FTS even in ‘asymptotic regime’ where Wilks’ applies

You *can* beat the likelihood ratio test statistic

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \quad\longrightarrow\quad FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

- Denominator in 'Focused Test Statistic' (FTS) knows about all alternate hypotheses
- $f(\mu)$ focuses statistical power in meaningful regions of parameter space
  - Particularly useful in small sample / small signal regime
- Fast critical value estimation with ML



Shorter median length for confidence intervals with FTS even in 'asymptotic regime' where Wilks' applies funding gets harder to secure, principal investigators are in their office writing grants while the trainees get to do the cool stuff.

Bryan W. Jones is a retinal neuroscientist at the University of Pittsburgh in Pennsylvania.

AISHIK GHOSH
STUDENTS OVERTURN LONG-HELD ASSUMPTION

I have worked on experimental particle physics since 2015, searching for Higgs bosons at CERN, Europe's particle-physics lab near Geneva, Switzerland, and now also working on the Deep Underground Neutrino Experiment (DUNE) in the United States. For this research, there's one statistical test we've used for decades to confirm the existence of a new particle – the generalized likelihood ratio test (GLRT). This compares two models – a simple null hypothesis, which includes no new particle or matter being discovered, and a more complicated alternative model, which includes a new particle with many possible values of strength.

In December 2024, a couple of PhD students working with my collaborator, Ann Lee, a data scientist at Carnegie Mellon University in Pittsburgh, Pennsylvania, were confident they could disprove the assumption that the GLRT was optimal. In the corner of my mind, I hoped they would prove us wrong. I gave them one of the most famous Higgs boson data sets to play around with. By early 2025, they showed that, although our previous physics results weren't wrong, our use of the GLRT wasn't ideal because it assumed large sample sizes are always generated, which is often not the case. Instead, the test left valuable information on the table. That day was special. I was still sceptical and I went through a battery of checks because I had to go back to my community and defend the PhD students' work, but it was all correct. The paper is currently in review, receiving a great deal of scrutiny.

Together, we produced a statistical test that will drastically improve our ability to make discoveries in particle physics, for example in searches for a new particle such as dark matter, where we expect to see only a few signal events at best. As a scientist, I want deeply held beliefs to be questioned. It was a real shock to the particle-physics community. Young people find it exciting. Senior members are still highly sceptical, as they should be, but they are coming around. As the DUNE experiment comes online, with this new statistical model in place, we hope to make precise measurements about neutrinos much sooner than anticipated.

Aishik Ghosh is a fundamental physicist at the Georgia Institute of Technology in Atlanta.

RAFIK TAREK NEME GARRIDO
SHOCKING CORAL FIND

A couple of years ago, after a day of pouring rain, the water on the Caribbean coast of Colombia was crystal clear and my master's student, Jorge Mareno, managed to take pictures of corals that no one knew existed here. We could find no scientific reports of corals in the area. Typically, the water is pretty turbid because the Magdalena River, which flows from the south of the country to the Caribbean Sea, brings chemicals and pollutants. It's an ongoing ecological and social challenge, but these corals must be adapting to these conditions. We did a sampling campaign across three days with a boat, using environmental DNA to find areas where corals, sponges and fish successfully survive the conditions. Most of the records are completely new for the region. It's super gratifying.

Rafik Tarek Neme Garrido is an evolutionary biologist at the University of the North in Barranquilla, Colombia.

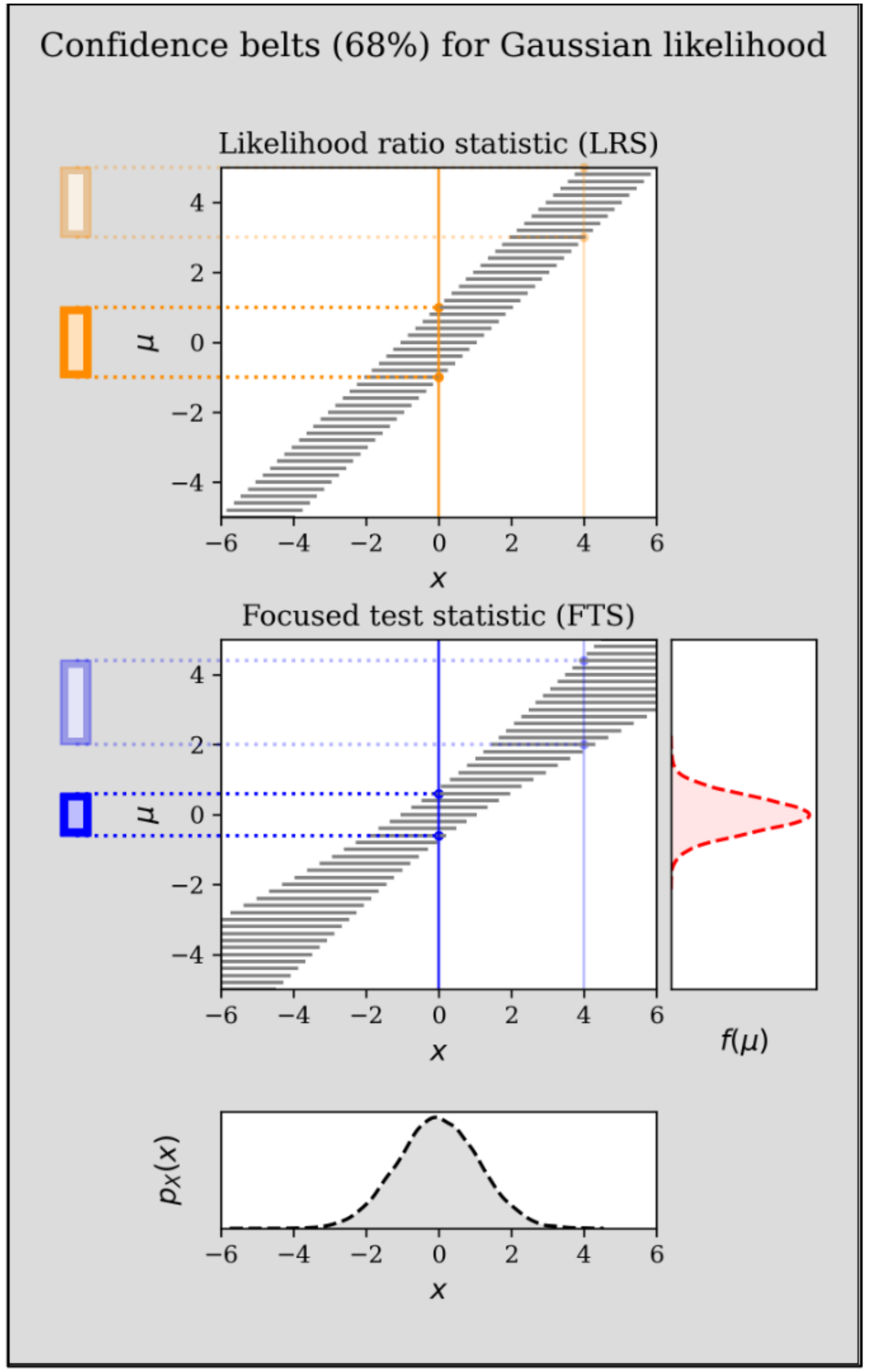
TIM CURRAN
BURN PREDICTIONS

In my group, we test the flammability of plant species using a barbecue. The results can help with fire-mitigation policies and with understanding the evolution of flammability. As part of an outreach activity, we host schoolchildren at the university who haven't had much exposure to academia before. We ask the kids to predict how a particular plant species will behave – for example, what characteristics will make it burn less or more – and then we see who is right. The kids get really into it. They ask amazing questions, the same kind that peer reviewers have asked us, including questioning our methodological assumptions, such as "why do you only blowtorch them for ten seconds?"

Most of the really good days doing science have been associated with young students having a light-bulb moment. In the rather

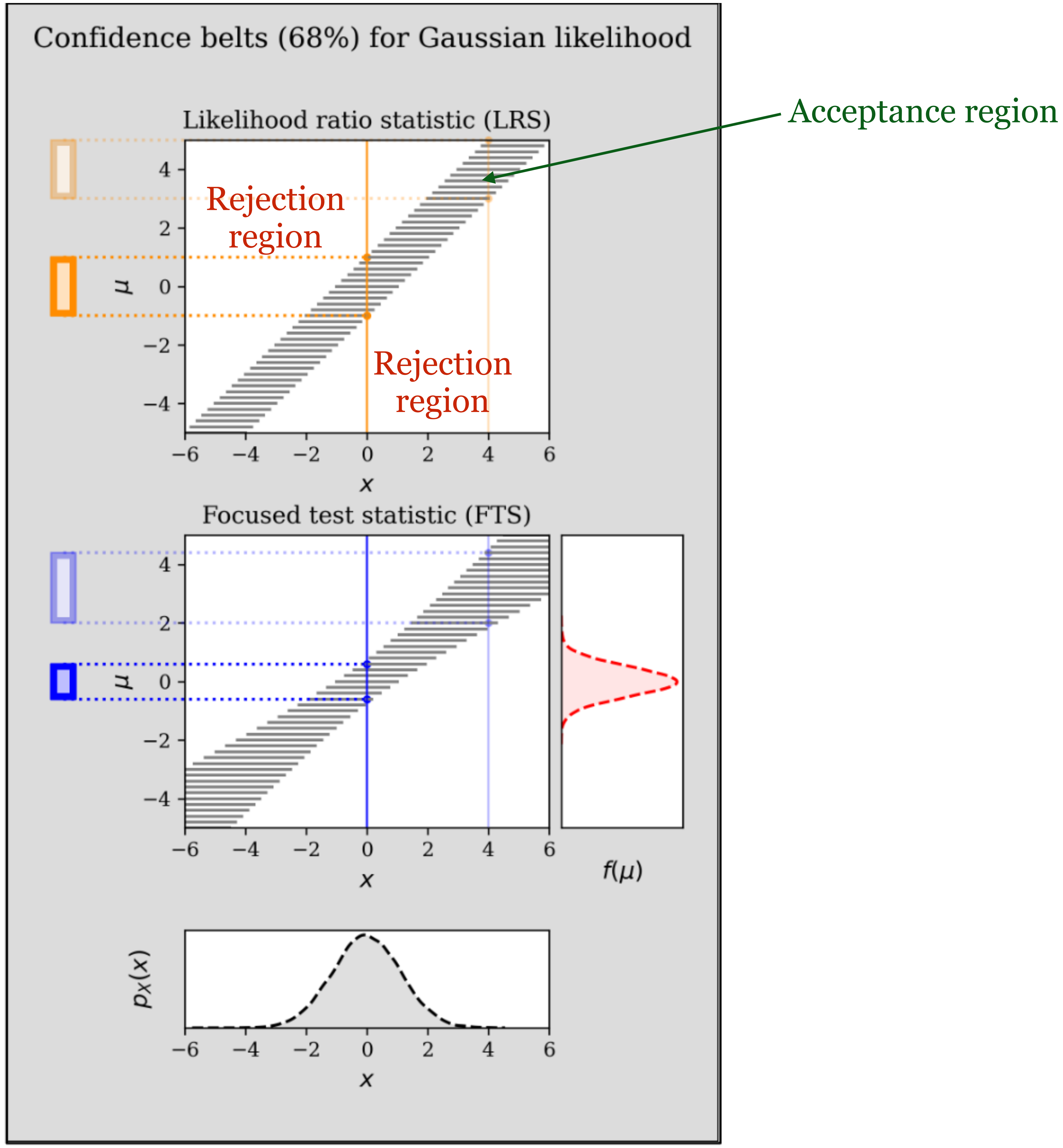
https://www.nature.com/articles/d41586-026-01479-8?utm_source=x&utm_medium=social&utm_campaign=nature&linkId=61827825

Test inversion for simplest measurement



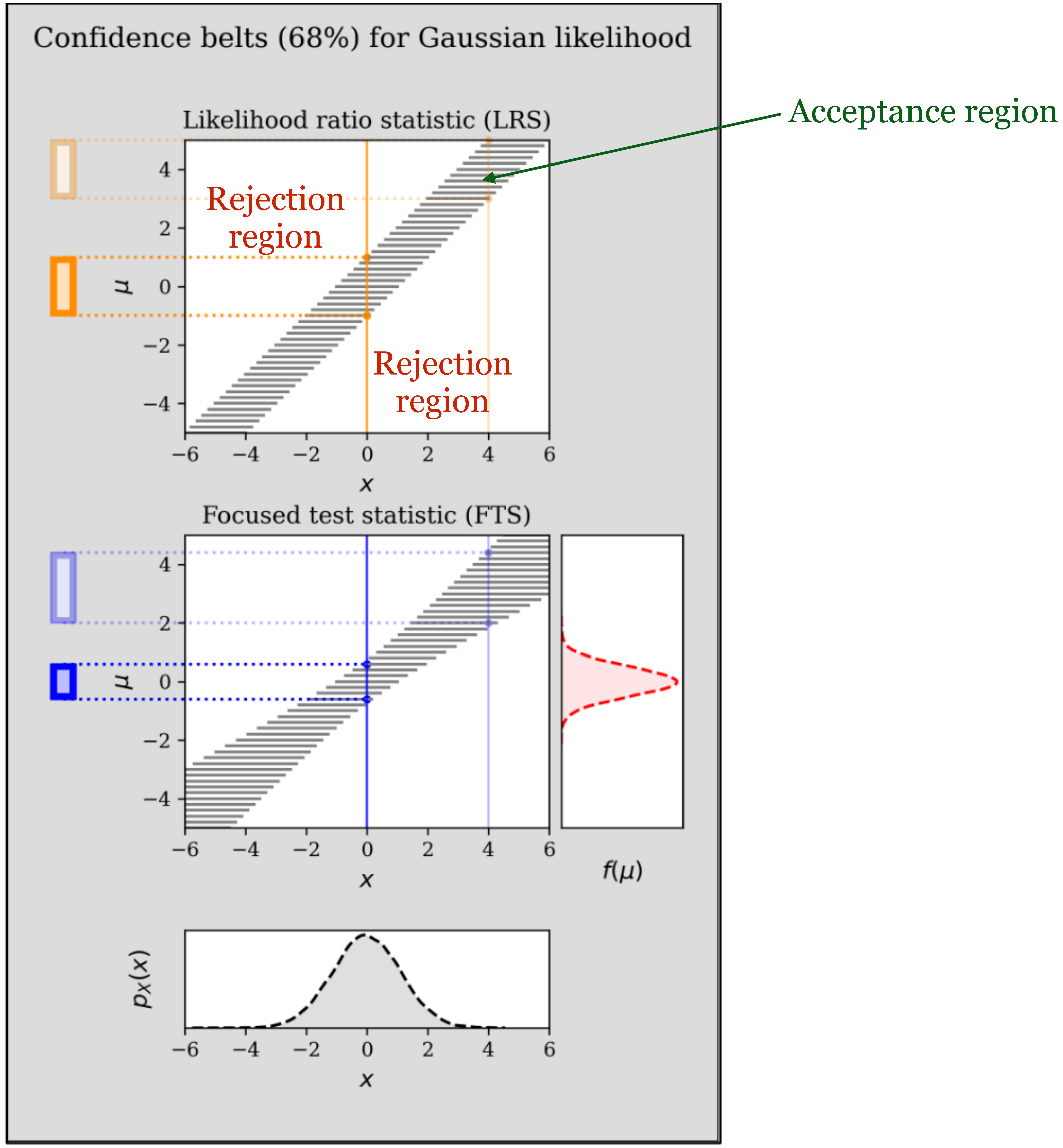
Data comprises single measurement x

Test inversion for simplest measurement



Data comprises single measurement x

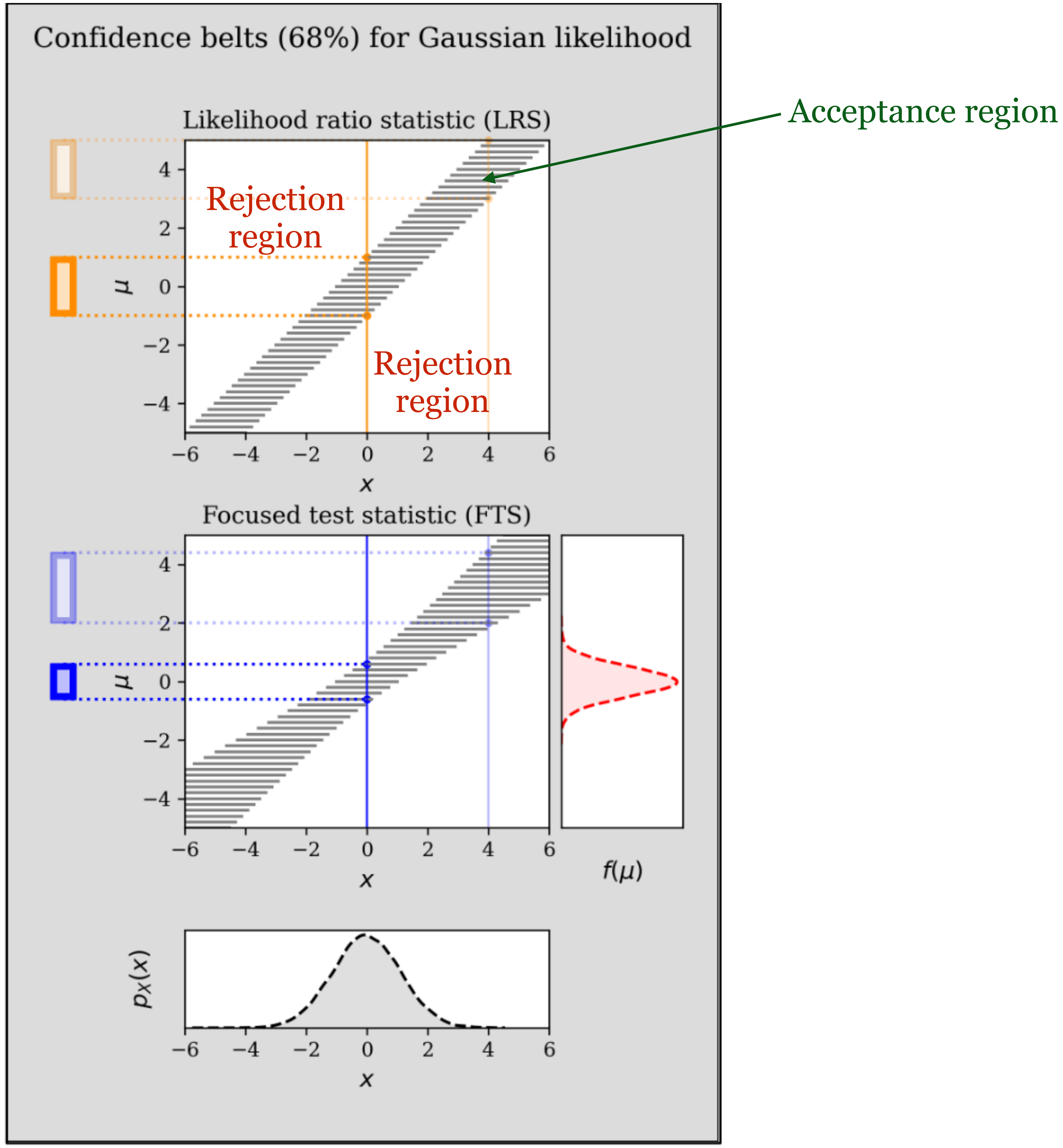
Test inversion for simplest measurement



Data comprises single measurement x

- LRS gives constant CI sizes
- FTS gives narrower near focus region by sacrificing power away from focus region

Test inversion for simplest measurement



Data comprises single measurement x

- LRS gives constant CI sizes
- FTS gives narrower near focus region by sacrificing power away from focus region
- You have full freedom to **choose any focus function** that gives you the **best expected sensitivity** in physics-motivated regions using simulated samples

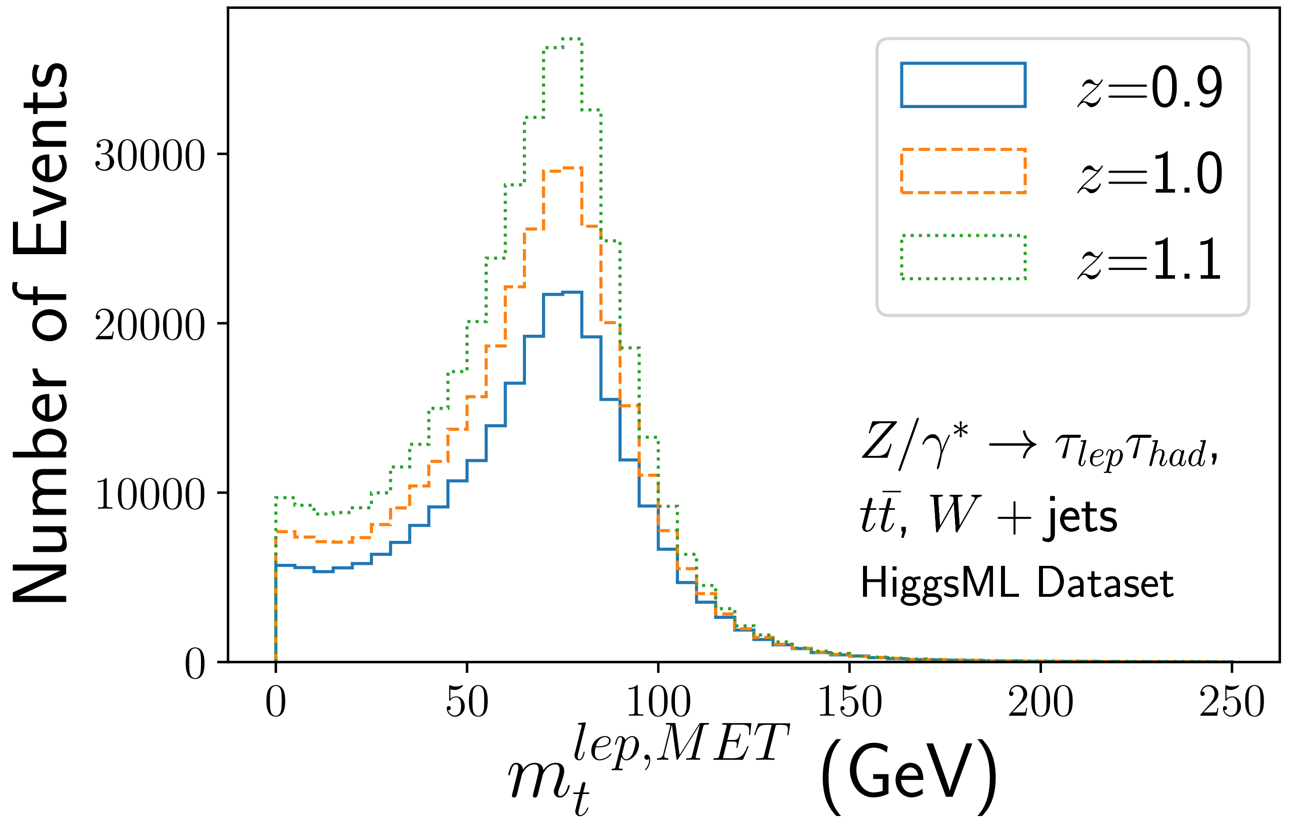
Open problems to extend to full ATLAS analysis:

- ✓ Robustness: Design and validation
- ‣ Systematic Uncertainties: Incorporate them in likelihood (ratio) model
- • Neyman Construction: Throwing toys in a per-event analysis

Systematic uncertainties

Experimental uncertainties:

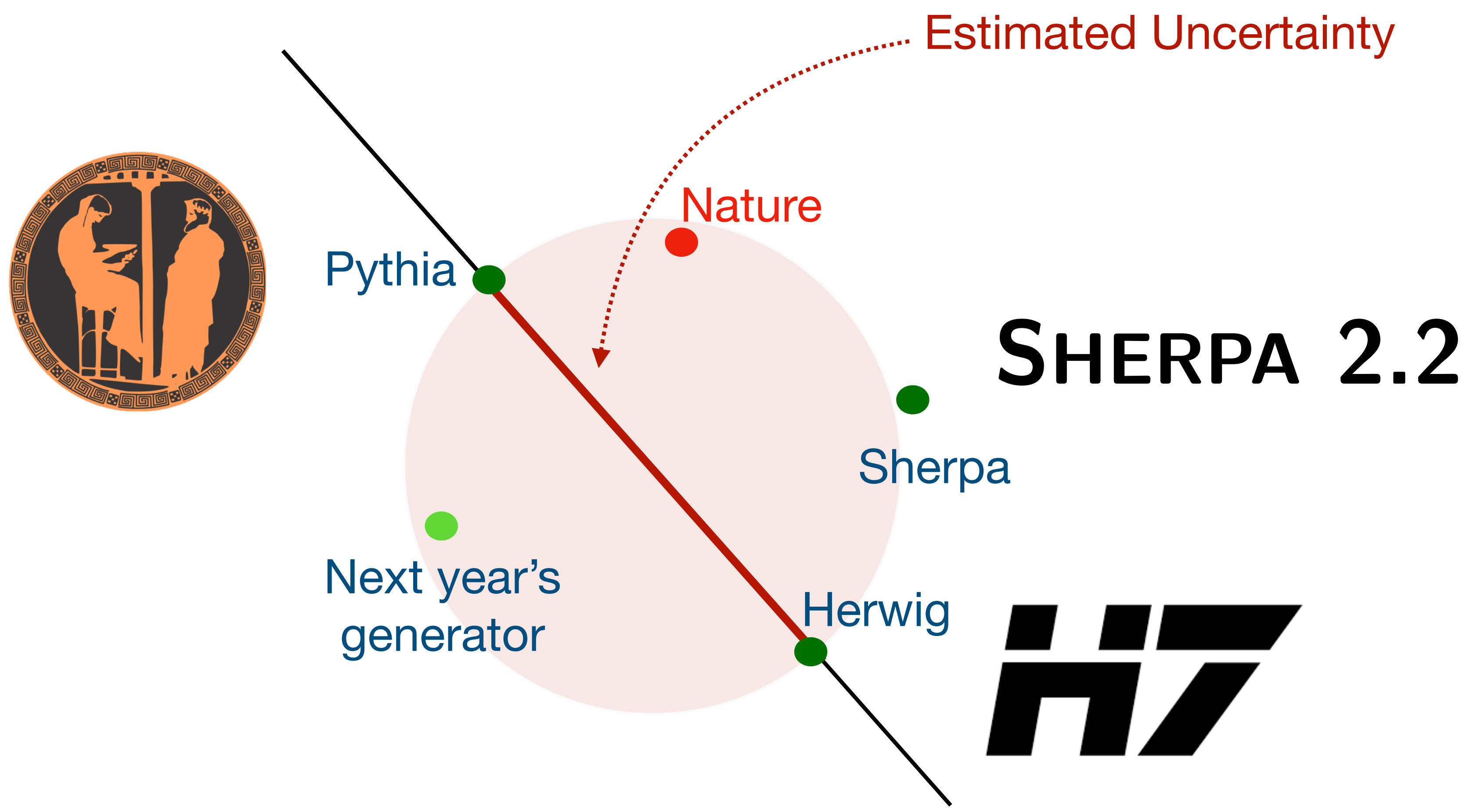
Eg. Inaccuracies in the calibration of our detector



Ghosh et al: PhysRevD.104.056026

Theory uncertainties:

Eg. Inability to compute QFT to infinite order



Ghosh and Nachman: epjc.s10052.022.10012.w

Systematic uncertainties

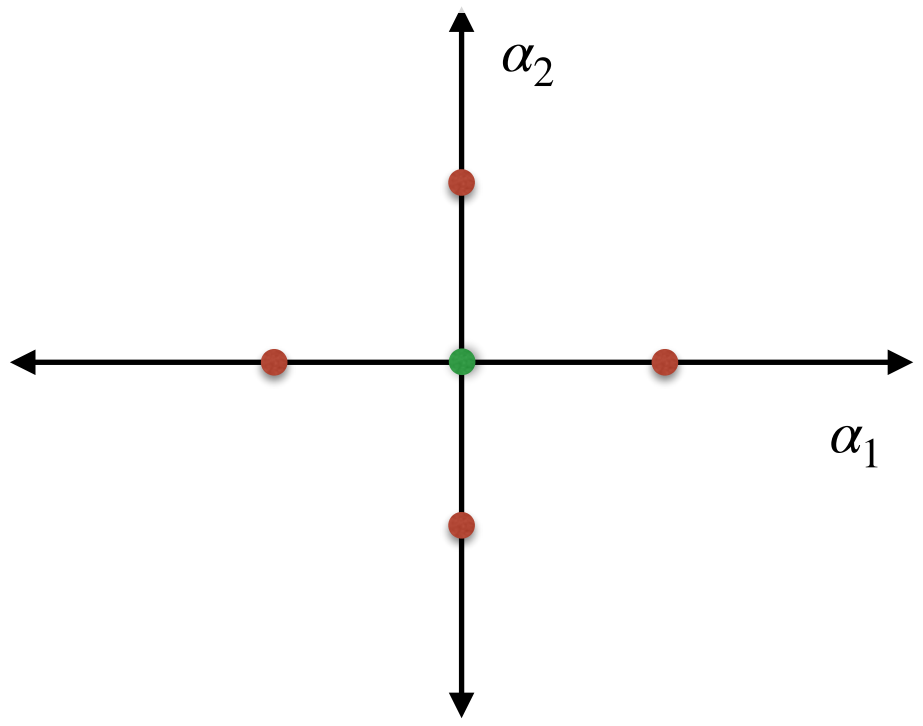
Experimental uncertainties:
Eg. Inaccuracies in the calibration of our detector

Ghosh et al: PhysRevD.104.056026

Theory uncertainties:
Eg. Inability to compute QFT to infinite order

Ghosh and Nachman: epjc.s10052.022.10012.w

- We only have simulations at 3 variations of each nuisance parameter α_k



Known interpolation strategies

See formula used in backup

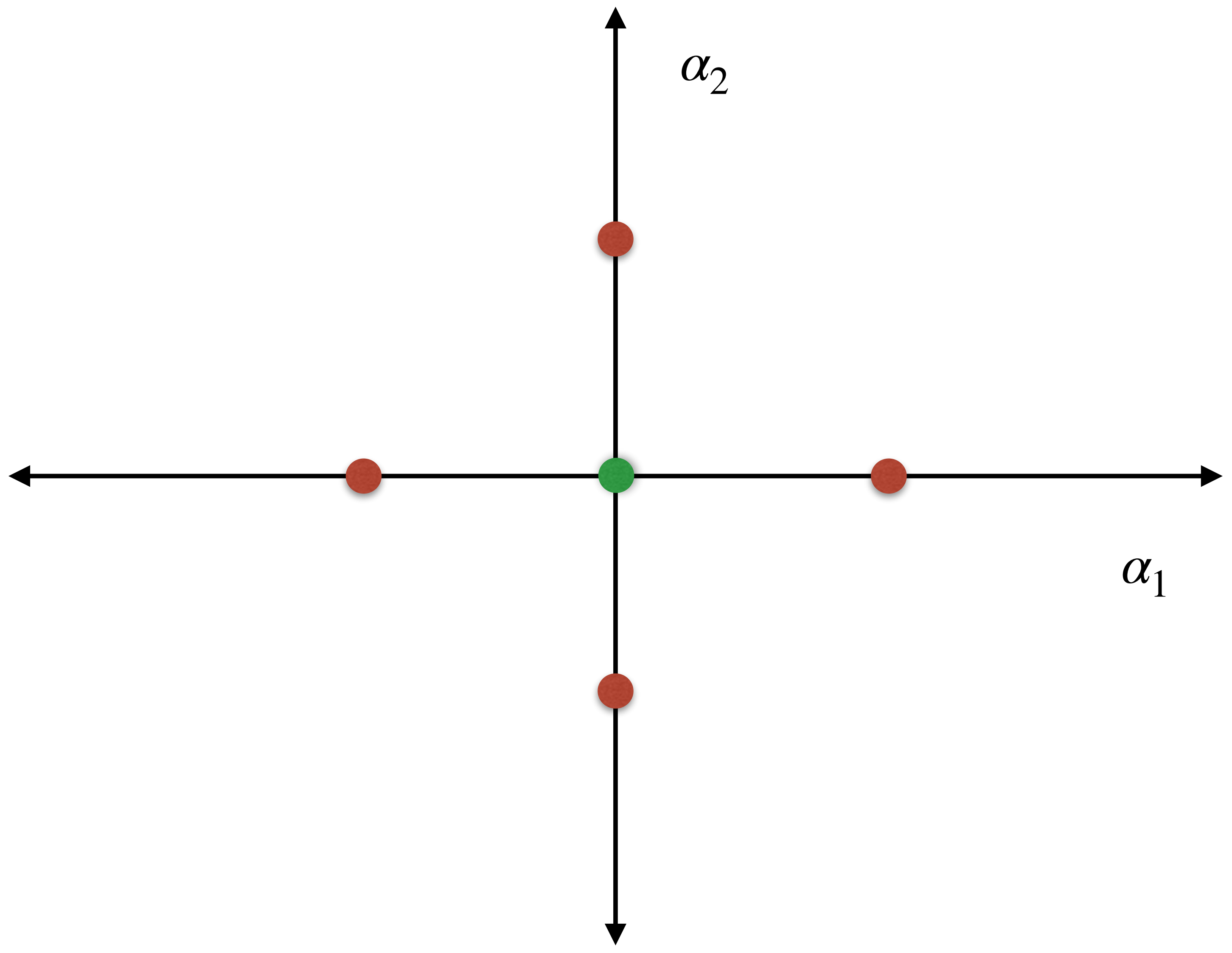
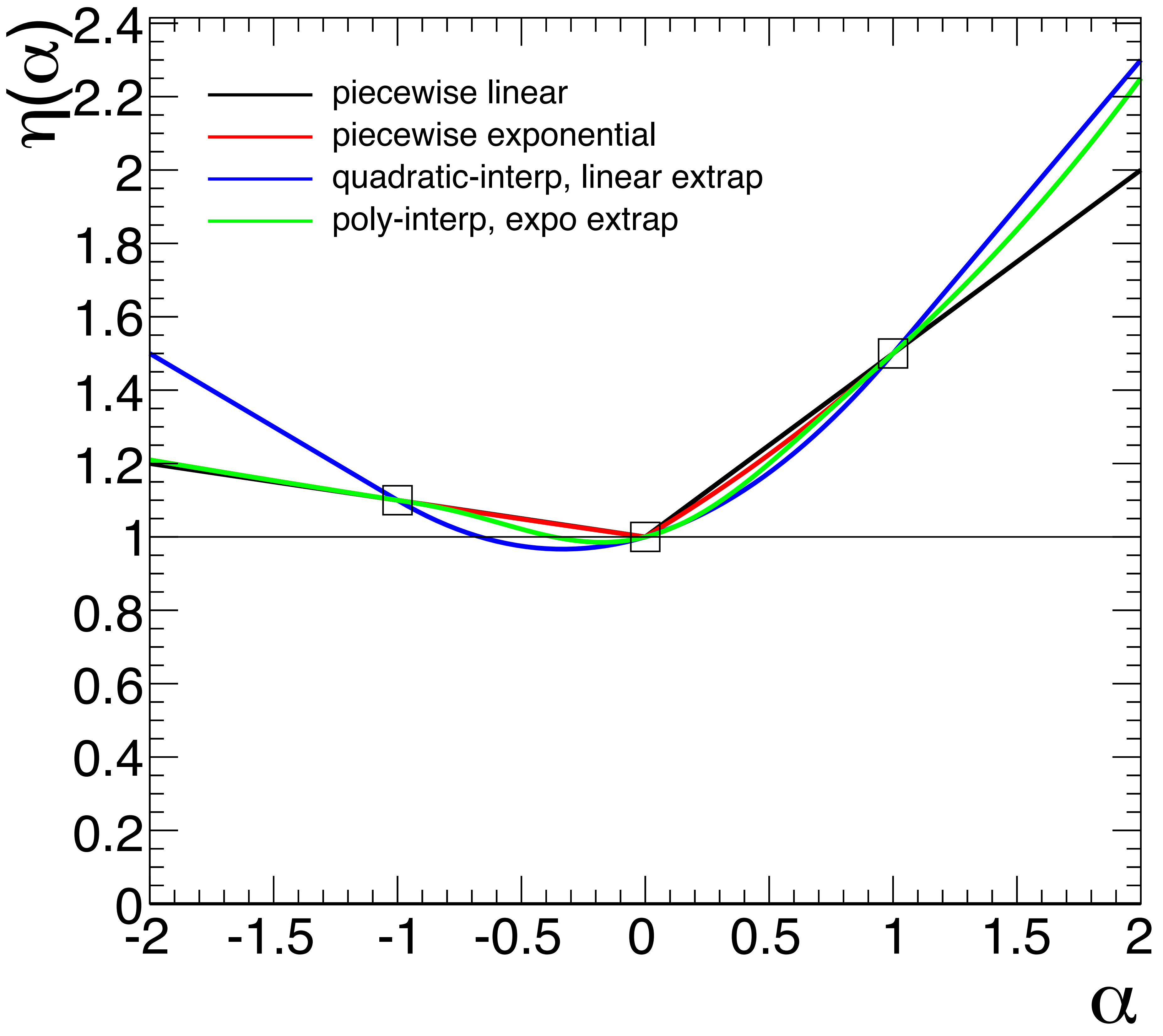


Image: arXiv:1503.07622

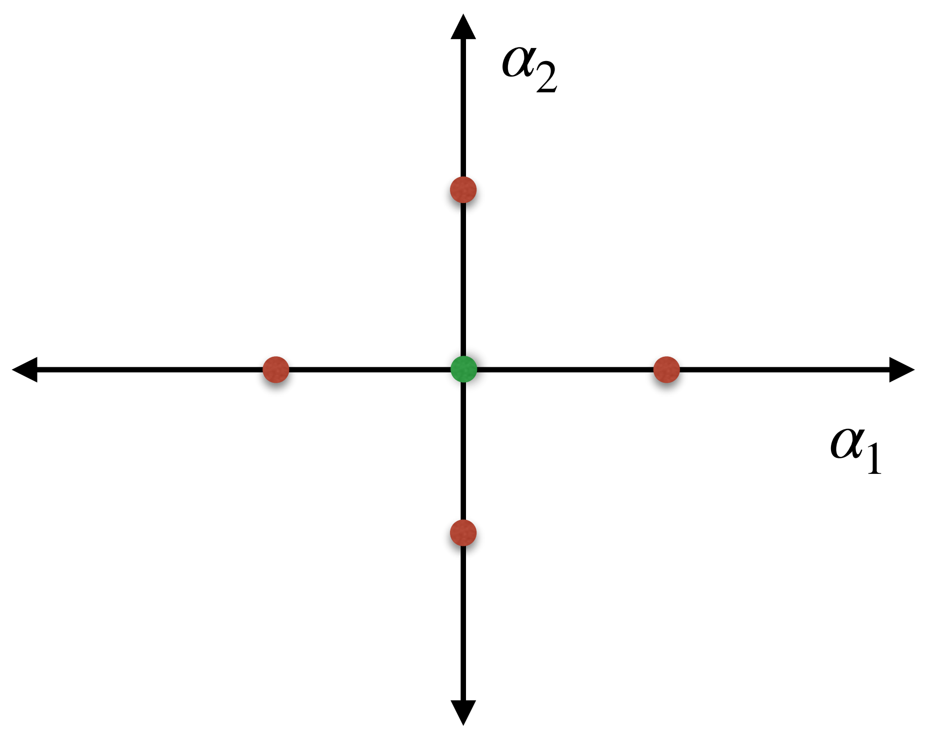


⇒ Combine these traditional interpolation with neural network estimation of per-event likelihood ratios

Probability density ratio including nuisance parameters (α)

x_i is one individual event

$$\frac{p(x_i \mid \mu, \alpha)}{p_{ref}(x_i)} =$$

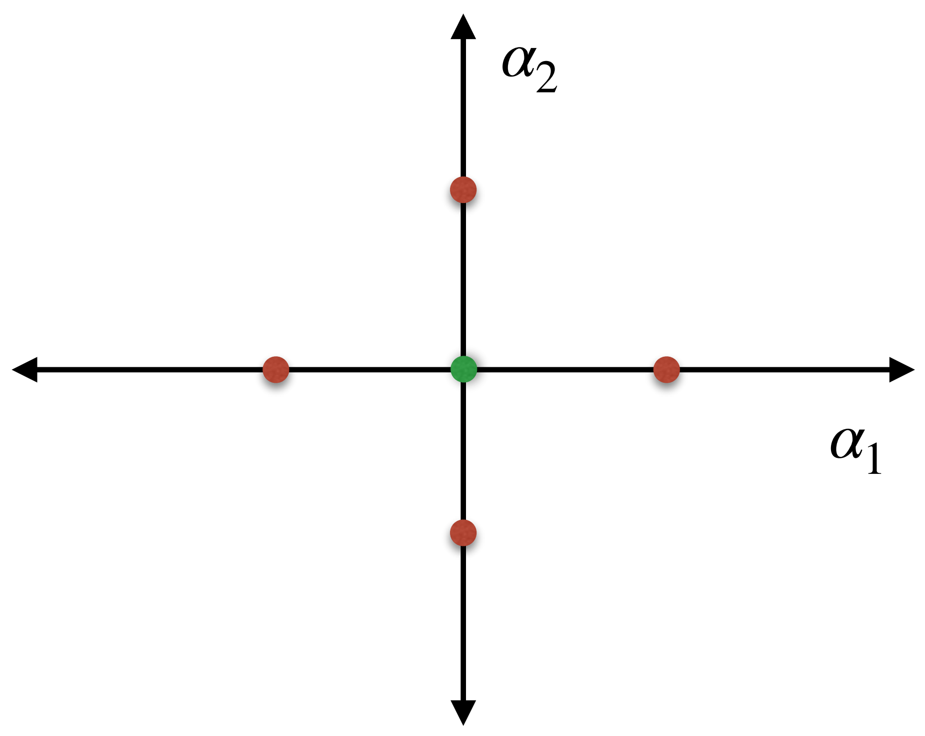


See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i, \alpha_k)$

Probability density ratio including nuisance parameters (α)

x_i is one individual event

$$\frac{p(x_i \,|\, \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu, \alpha)} \sum_j^C f_j(\mu) \cdot \nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_k^{N_{syst}} G_j(\alpha_k) \cdot g_j(x_i, \alpha_k)$$



$$g_j(x_i, \alpha_k) = \frac{p_j(x_i, \alpha_k)}{p_j(x_i)}$$

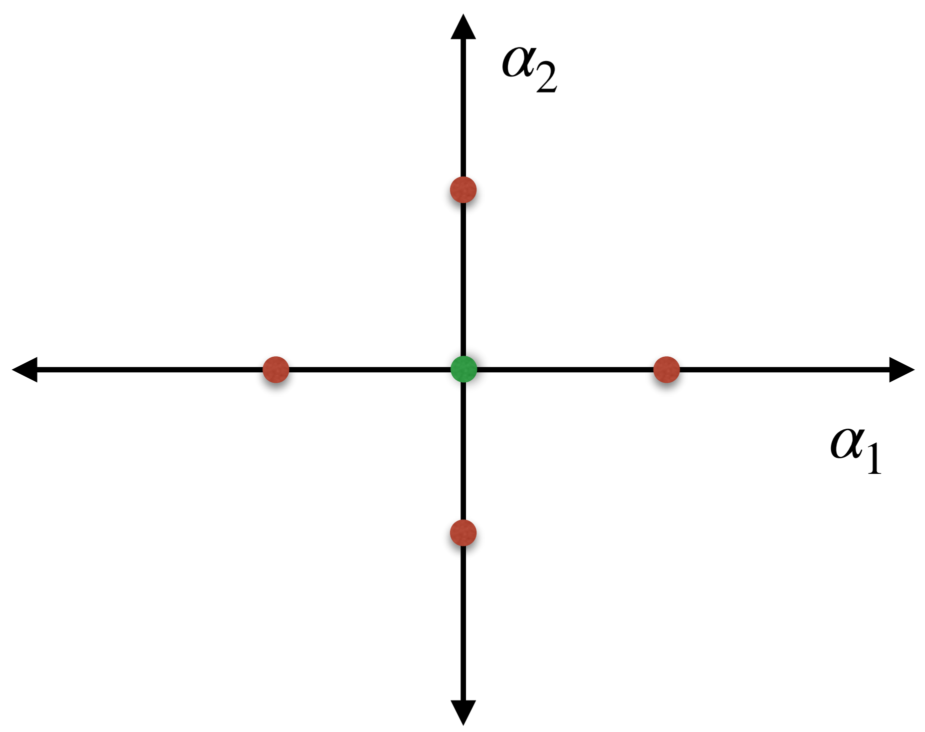
See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i, \alpha_k)$

Probability density ratio including nuisance parameters (α)

x_i is one individual event

$$\frac{p(x_i\,|\,\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu,\alpha)} \sum_j^C f_j(\mu)\cdot\nu_j\cdot\frac{p_j(x_i)}{p_{ref}(x_i)}\cdot\prod_k^{N_{syst}} G_j(\alpha_k)\cdot g_j(x_i,\alpha_k)$$

We have this already



$$g_j(x_i,\alpha_k) = \frac{p_j(x_i,\alpha_k)}{p_j(x_i)}$$

See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i,\alpha_k)$

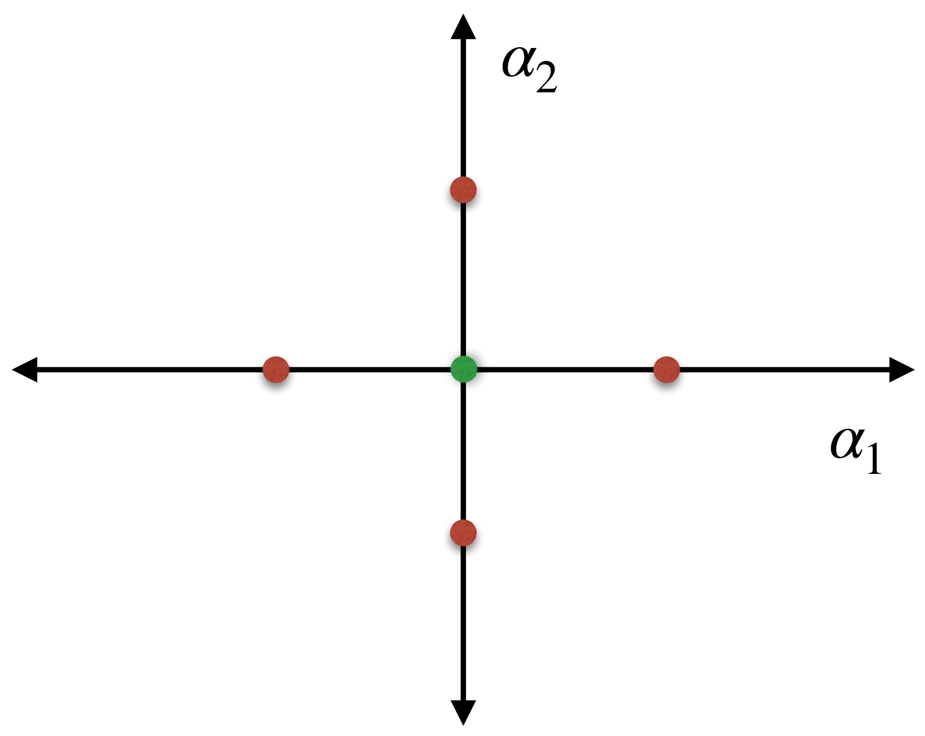
Probability density ratio including nuisance parameters (α)

x_i is one individual event

$$\frac{p(x_i \mid \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu, \alpha)} \sum_j^C f_j(\mu) \cdot \nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_k^{N_{syst}} G_j(\alpha_k) \cdot g_j(x_i, \alpha_k)$$

We have this already

Estimate from simulations and existing interpolation methods



$$g_j(x_i, \alpha_k) = \frac{p_j(x_i, \alpha_k)}{p_j(x_i)}$$

See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i, \alpha_k)$

Probability density ratio including nuisance parameters (α)

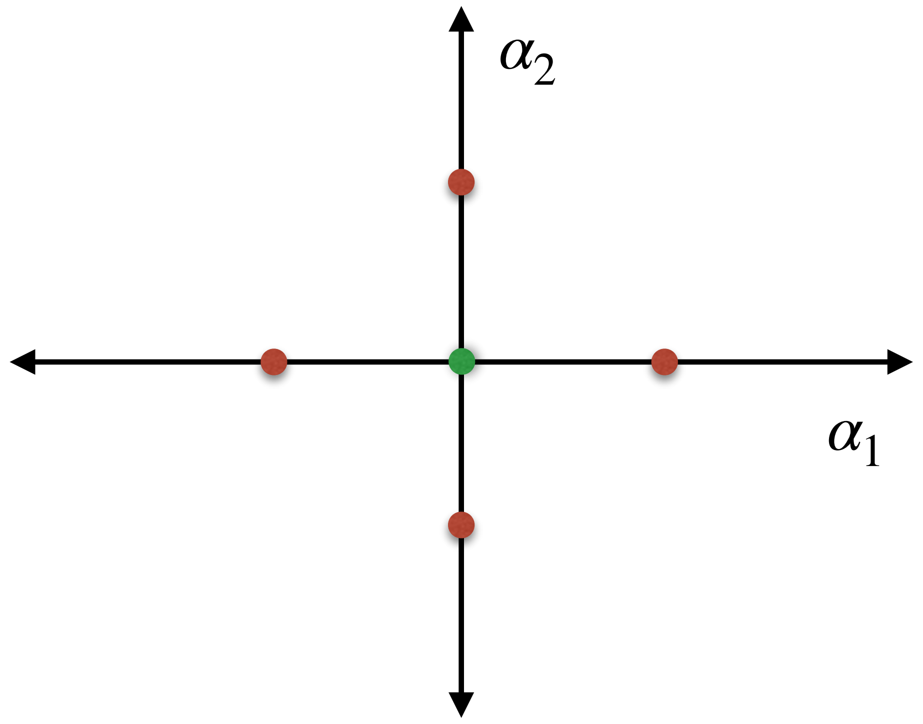
x_i is one individual event

$$\frac{p(x_i \,|\, \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu, \alpha)} \sum_j^{C} f_j(\mu) \cdot \nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_k^{N_{syst}} G_j(\alpha_k) \cdot g_j(x_i, \alpha_k)$$

We have this already

Per-event terms estimated using another ensemble of networks and interpolation methods

$$g_j(x_i, \alpha_k) = \frac{p_j(x_i, \alpha_k)}{p_j(x_i)}$$



Estimate from simulations and existing interpolation methods

See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i, \alpha_k)$

Final test statistic

x_i is one individual event

$$\frac{L_{\text{full}}(\mu, \alpha | \mathcal{D})}{L_{\text{ref}}(\mathcal{D})} = \text{Pois}(N_{\text{data}} | \nu(\mu, \alpha)) \prod_{i}^{N_{\text{data}}} \frac{p(x_i | \mu, \alpha)}{p_{\text{ref}}(x_i)} \prod_{k} \text{Gaus}(a_k | \alpha_k, \delta_k)$$

Final test statistic

x_i is one individual event

From previous slide

$$\frac{L_{\text{full}}(\mu, \alpha | \mathcal{D})}{L_{\text{ref}}(\mathcal{D})} = \text{Pois}(N_{\text{data}} | \nu(\mu, \alpha)) \prod_i^{N_{\text{data}}} \frac{p(x_i | \mu, \alpha)}{p_{\text{ref}}(x_i)} \prod_k \text{Gaus}(a_k | \alpha_k, \delta_k)$$

Final test statistic

x_i is one individual event

From previous slide

$$\frac{L_{\text{full}}(\mu, \alpha | \mathcal{D})}{L_{\text{ref}}(\mathcal{D})} = \text{Pois}(N_{\text{data}} | \nu(\mu, \alpha)) \prod_{i}^{N_{\text{data}}} \frac{p(x_i | \mu, \alpha)}{p_{\text{ref}}(x_i)} \prod_{k} \text{Gaus}(a_k | \alpha_k, \delta_k)$$

Prod over events

ATLAS methods paper: arXiv:2412.01600

Final test statistic

x_i is one individual event

$$\frac{L_{\text{full}}(\mu, \alpha|\mathcal{D})}{L_{\text{ref}}(\mathcal{D})} = \text{Pois}(N_{\text{data}}|\nu(\mu, \alpha)) \prod_{i}^{N_{\text{data}}} \frac{p(x_i|\mu, \alpha)}{p_{\text{ref}}(x_i)} \prod_{k} \text{Gaus}(a_k|\alpha_k, \delta_k)$$

From previous slide

Rate term

Prod over events

Final test statistic

x_i is one individual event

$$\frac{L_{\text{full}}(\mu,\alpha|\mathcal{D})}{L_{\text{ref}}(\mathcal{D})} = \text{Pois}(N_{\text{data}}|\nu(\mu,\alpha)) \prod_{i}^{N_{\text{data}}} \frac{p(x_i|\mu,\alpha)}{p_{\text{ref}}(x_i)} \prod_{k} \text{Gaus}(a_k|\alpha_k,\delta_k)$$

From previous slide

Rate term

Prod over events

Constrain term

Final test statistic

x_i is one individual event

$$\frac{L_{\text{full}}(\mu,\alpha|\mathcal{D})}{L_{\text{ref}}(\mathcal{D})} = \text{Pois}(N_{\text{data}}|\nu(\mu,\alpha)) \prod_i^{N_{\text{data}}} \frac{p(x_i|\mu,\alpha)}{p_{\text{ref}}(x_i)} \prod_k \text{Gaus}(a_k|\alpha_k,\delta_k)$$

From previous slide

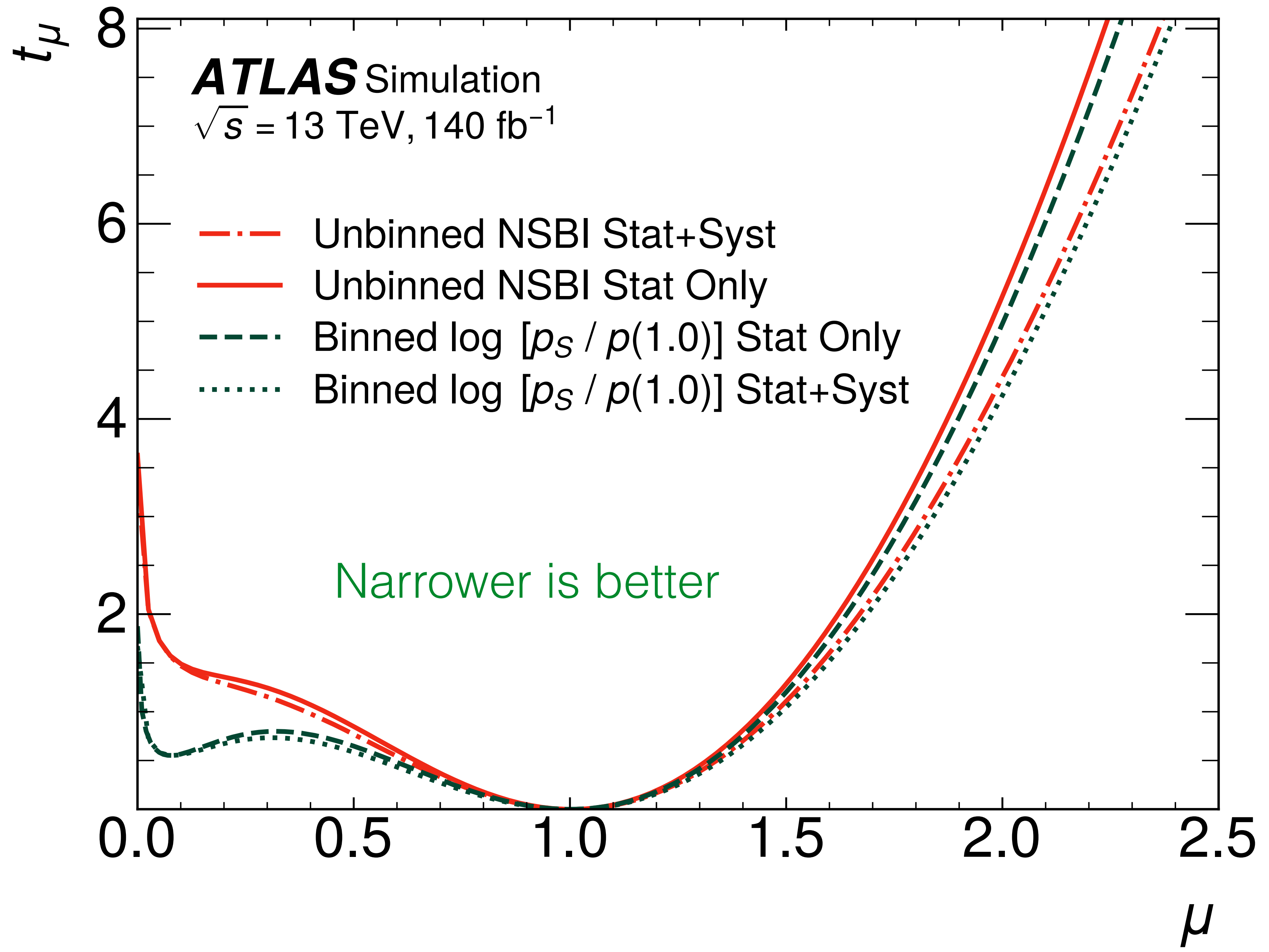
Rate term

Prod over events

Constrain term

Profiling: $$t_\mu = -2\ln\left(\frac{L_{\text{full}}(\mu,\hat{\hat{\alpha}})/L_{\text{ref}}}{L_{\text{full}}(\hat{\mu},\hat{\alpha})/L_{\text{ref}}}\right)$$

This is why we define p_{ref} to be independent of μ



Final test statistic

x_i is one individual event

$$\frac{L_{\text{full}}(\mu,\alpha|\mathcal{D})}{L_{\text{ref}}(\mathcal{D})} = \text{Pois}(N_{\text{data}}|\nu(\mu,\alpha)) \prod_{i}^{N_{\text{data}}} \frac{p(x_i|\mu,\alpha)}{p_{\text{ref}}(x_i)} \prod_k \text{Gaus}(a_k|\alpha_k,\delta_k)$$

From previous slide

Rate term

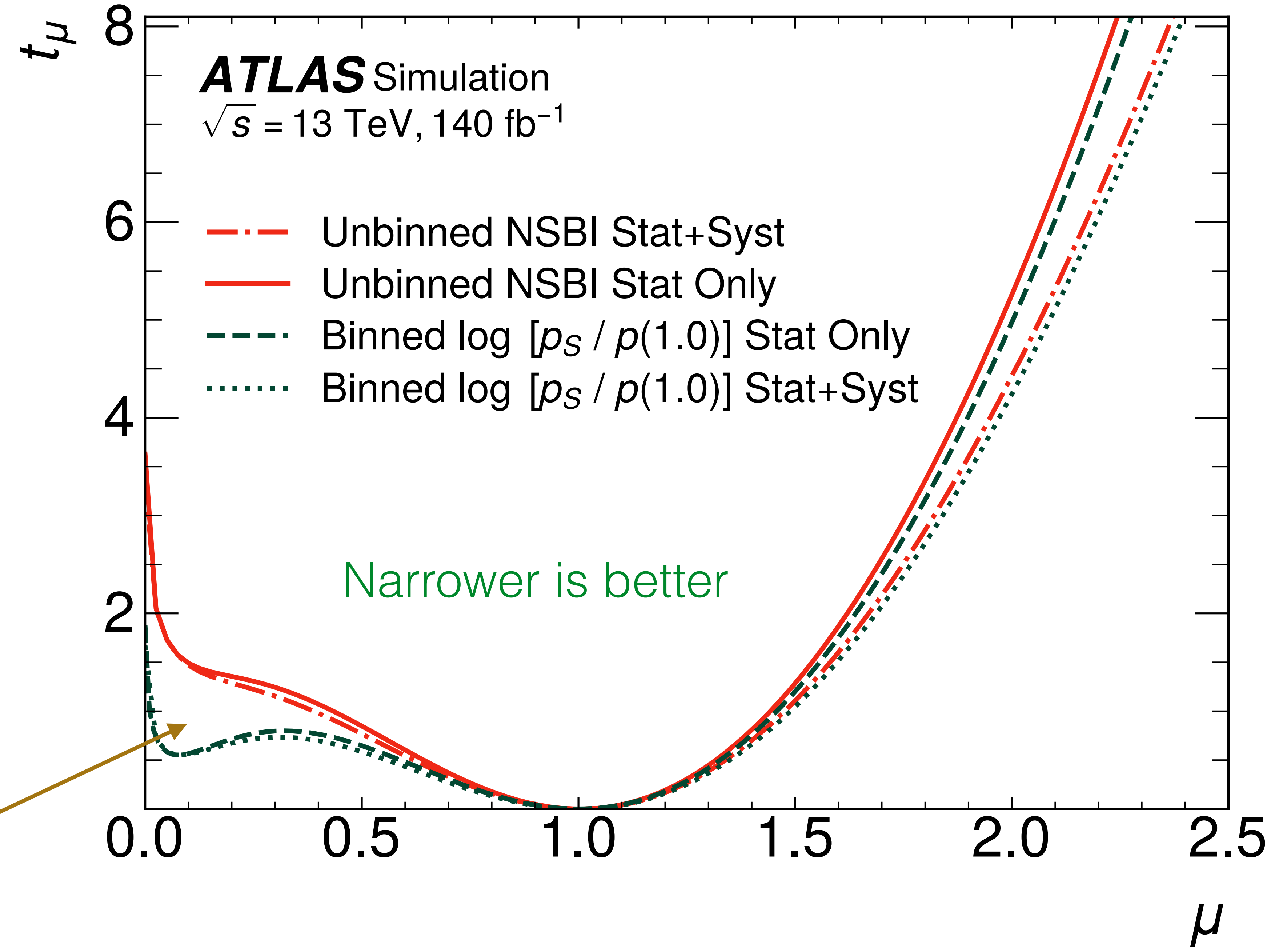
Prod over events

Constrain term

Profiling:

$$t_\mu = -2\ln\left(\frac{L_{\text{full}}(\mu,\hat{\hat{\alpha}})/\cancel{L_{\text{ref}}}}{L_{\text{full}}(\hat{\mu},\hat{\alpha})/\cancel{L_{\text{ref}}}}\right)$$

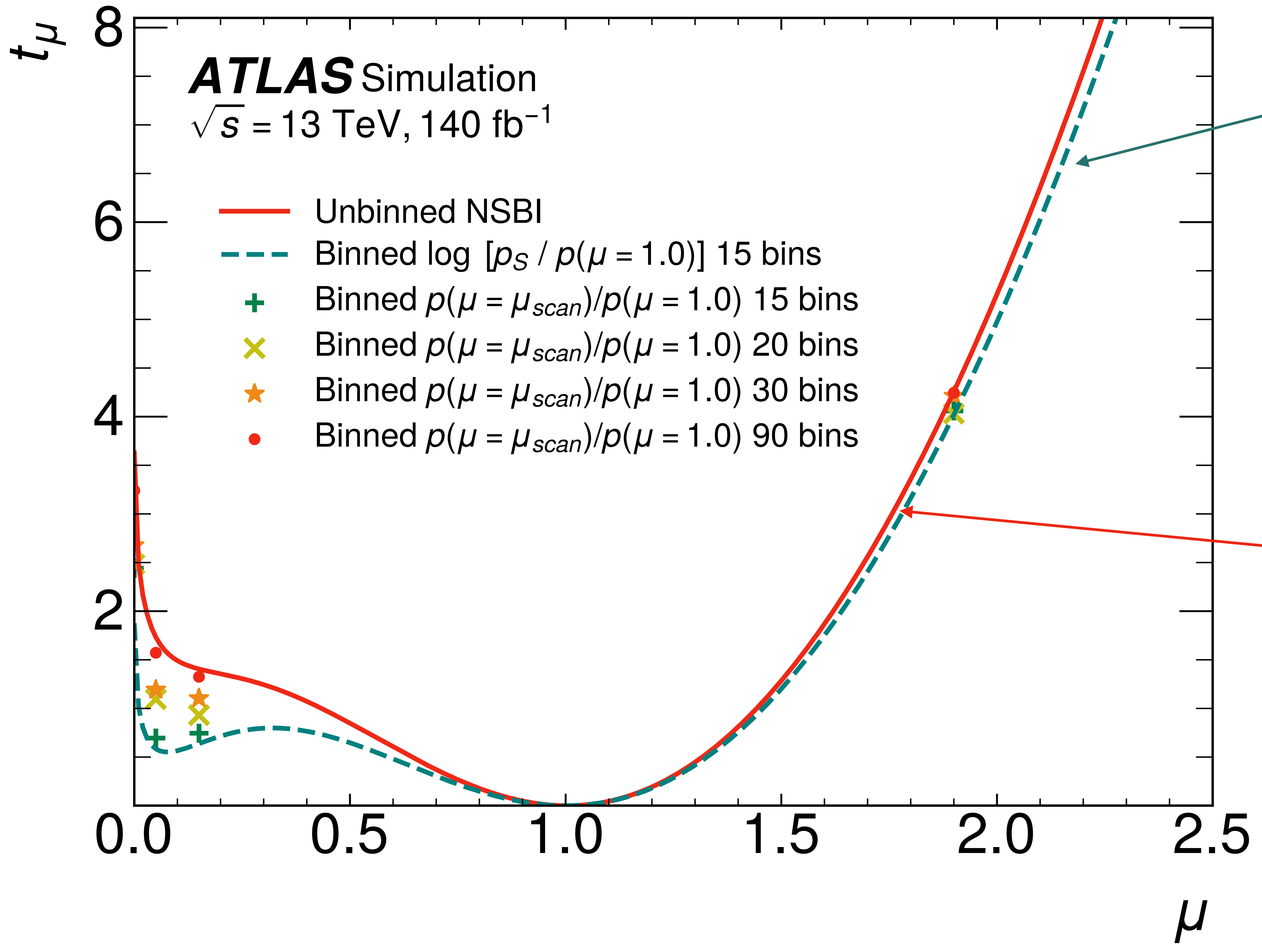
This is why we define p_{ref} to be independent of μ



Non-parabolic shape due to non-linear effects from quantum interference

Why does NSBI work better than traditional analyses?

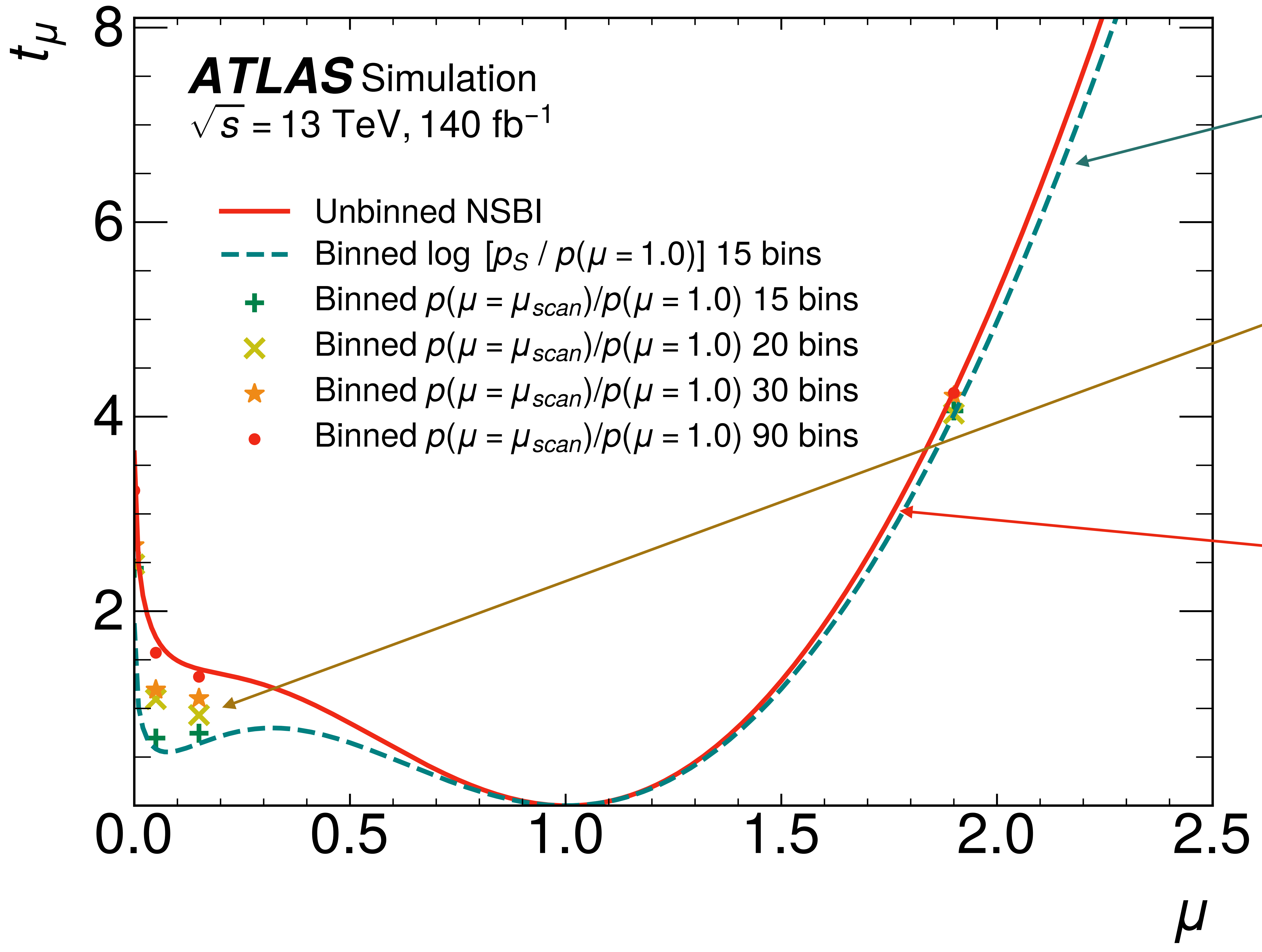
Why does it work better than traditional analyses?



$O_{fixed} = \log \frac{p_S(x_i)}{p_{SBI}(x_i)}$: Similar to histogram analysis

NSBI: Optimised for all hypotheses, unbinned

Why does it work better than traditional analyses?

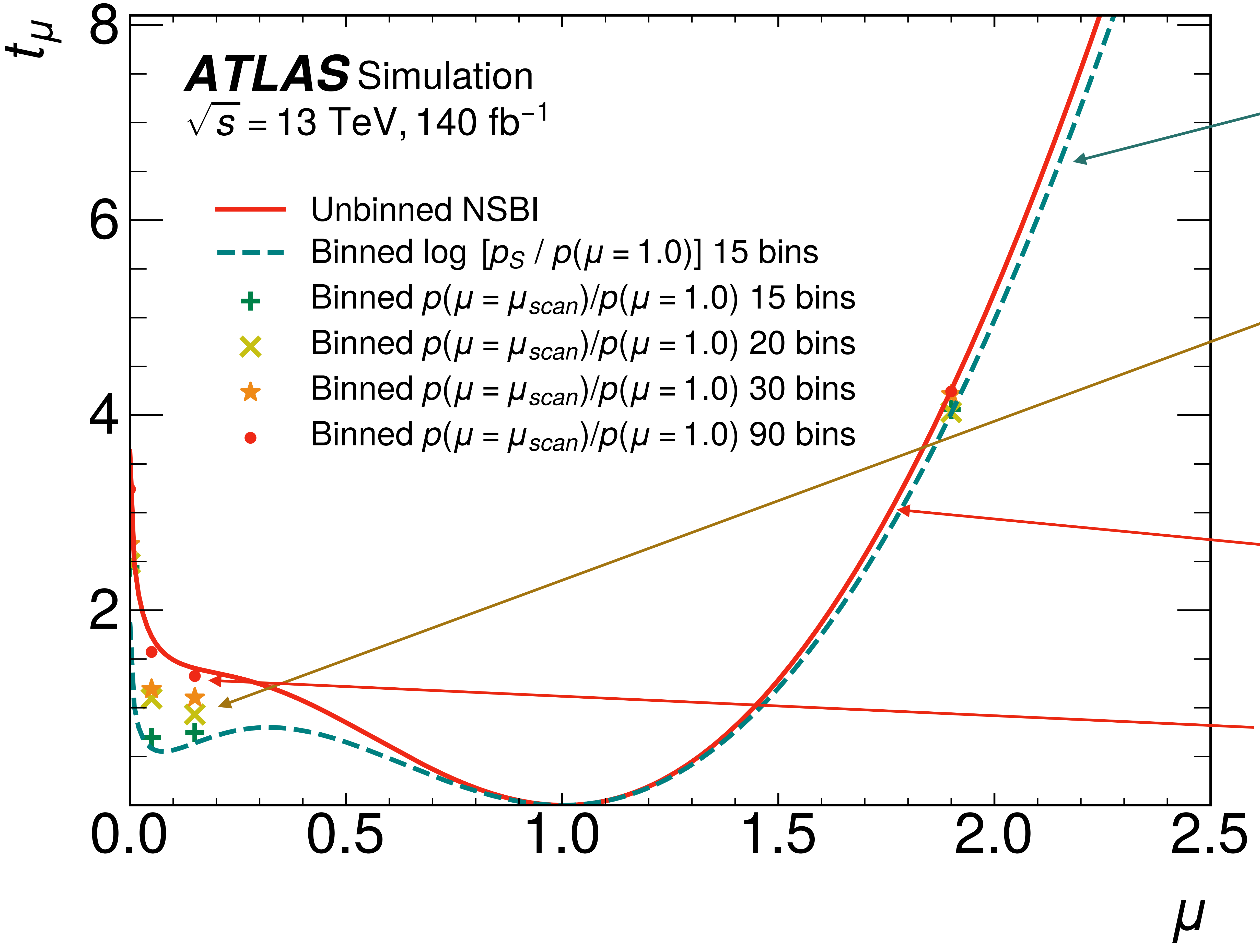


$O_{\text{fixed}} = \log \frac{p_S(x_i)}{p_{SBI}(x_i)}$: Similar to histogram analysis

$O_\mu = \frac{p(x_i|\mu)}{p(x_i|\mu = 1)}$: Hypothesis-specific observable, histogram fit

NSBI: Optimised for all hypotheses, unbinned

Why does it work better than traditional analyses?



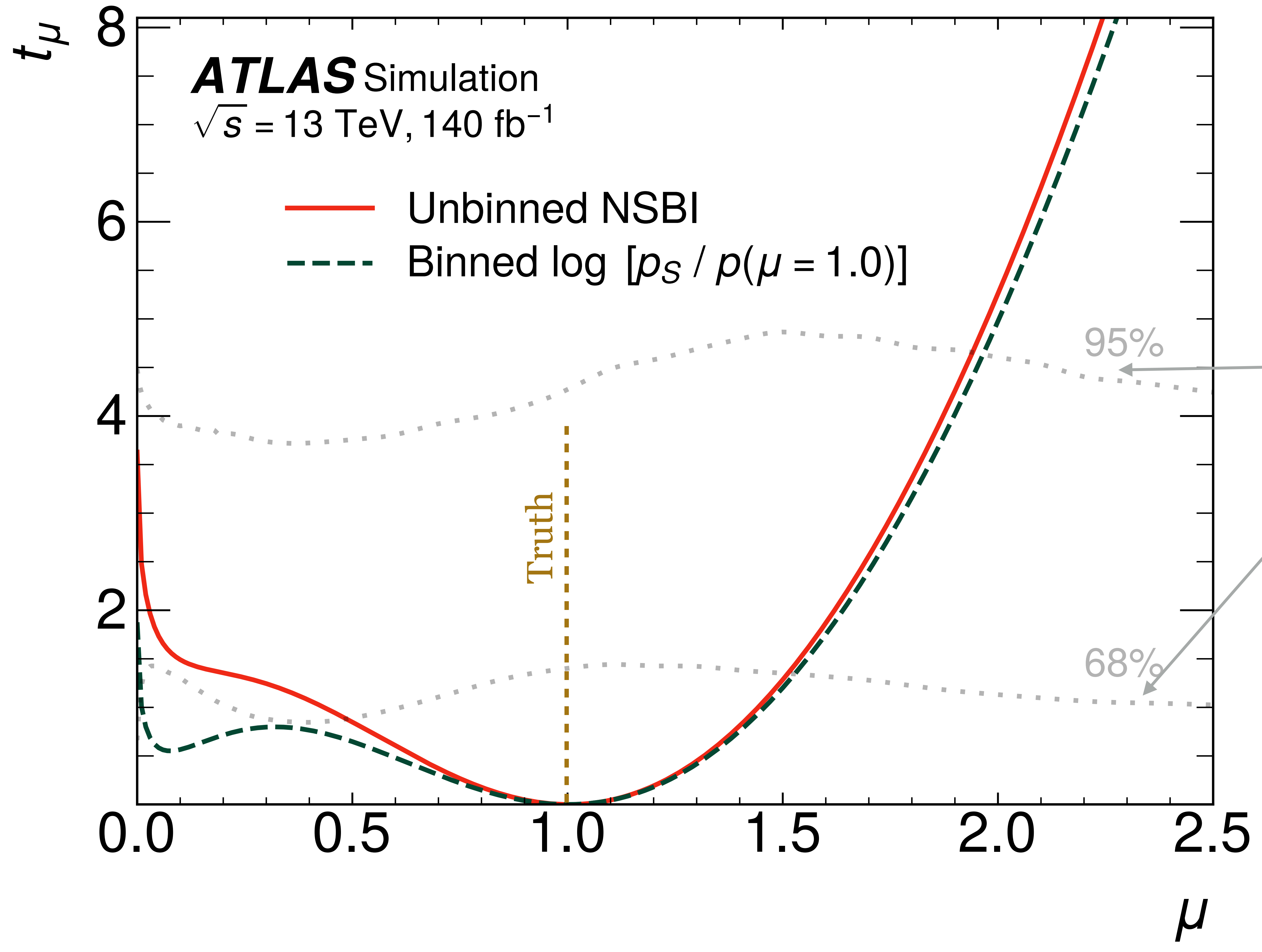
$O_{fixed} = \log \frac{p_S(x_i)}{p_{SBI}(x_i)}$: Similar to histogram analysis

$O_\mu = \frac{p(x_i | \mu)}{p(x_i | \mu = 1)}$: Hypothesis-specific observable, histogram fit

NSBI: Optimised for all hypotheses, unbinned

O_μ approaches NSBI as nBins $\to \infty$

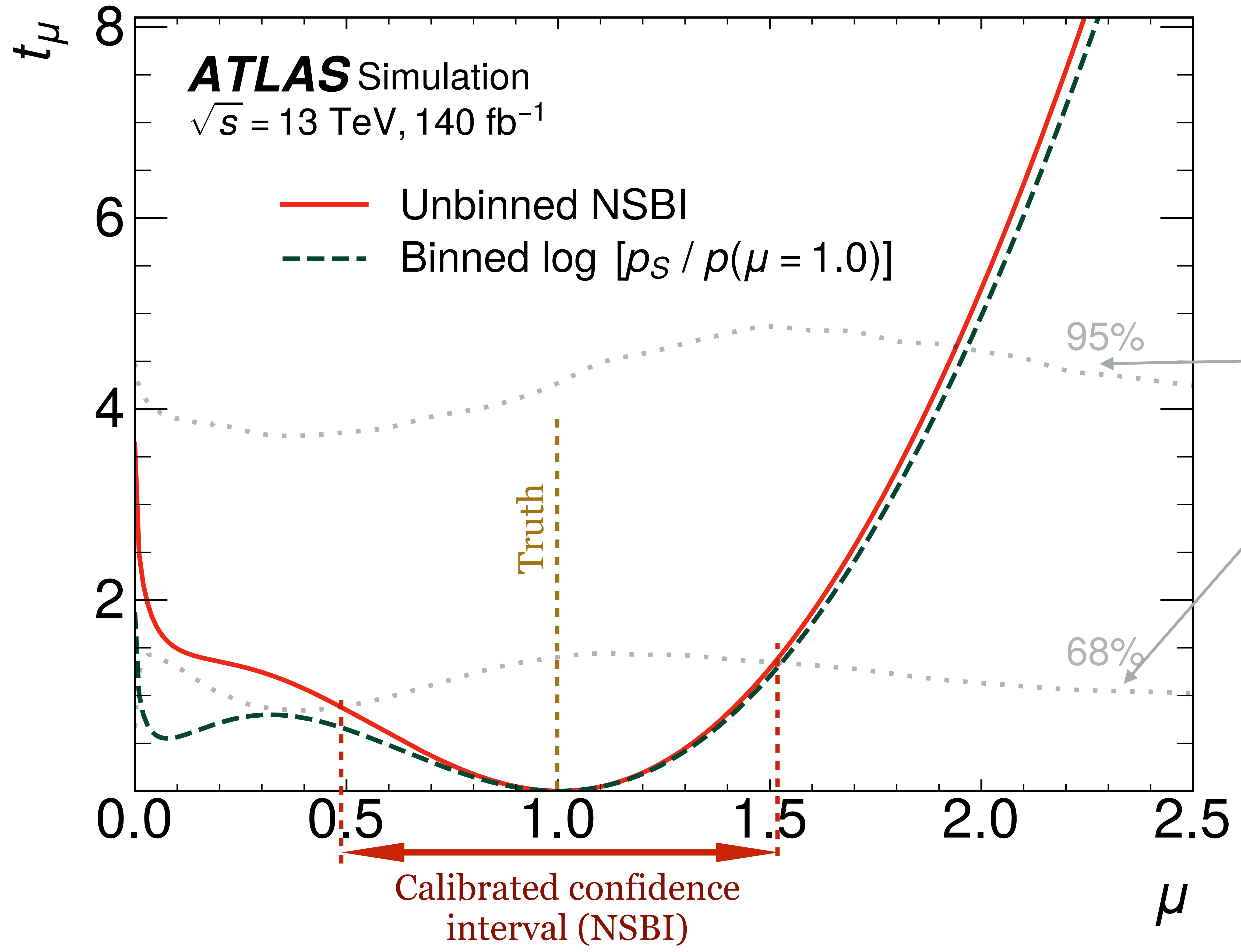
Critical values



Built with Neyman construction: Expensive process of running entire analysis on thousands of pseudo-experiments

Similar to structure seen in histogram analysis

Critical values



Built with Neyman construction: Expensive process of running entire analysis on thousands of pseudo-experiments

Similar to structure seen in histogram analysis

Vertical interpolation

$$G_j(\alpha_k) = \begin{cases} \left(\frac{\nu_j(\alpha_k^+)}{\nu_j(\alpha_k^0)}\right)^{\alpha_k} & \alpha_k > 1 \\ 1 + \sum_{n=1}^{6} c_n \alpha_k^n & -1 \leq \alpha_k \leq 1 \\ \left(\frac{\nu_j(\alpha_k^-)}{\nu_j(\alpha_k^0)}\right)^{-\alpha_k} & \alpha_k < -1 \end{cases}$$

$$g_j(x_i, \alpha_k) = \begin{cases} \left(g_j(x_i, \alpha_k^+)\right)^{\alpha_k} & \alpha_k > 1 \\ 1 + \sum_{n=1}^{6} c_n \alpha_k^n & -1 \leq \alpha_k \leq 1 \\ \left(g_j(x_i, \alpha_k^-)\right)^{-\alpha_k} & \alpha_k < -1 \end{cases}$$

With some continuity requirements