An implementation of Neural Simulation-Based Inference for Parameter Estimation in ATLAS

IN2P3/IRFU ML Workshop, Strasbourg
22 November 2024

Arnaud Maury, (David Rousseau), on behalf of the ATLAS Collaboration

Run 2 analysis of the off-shell Higgs boson decaying into four leptons

1 analysis, 2 papers:

- A Physics measurement conf note (to be submitted for publication imminently): https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2024-016/

- An ML-focused methodology paper (to be submitted for publication imminently) (this talk): https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2024-015/

The motivation for Neural Simulation-Based Inference (NSBI)

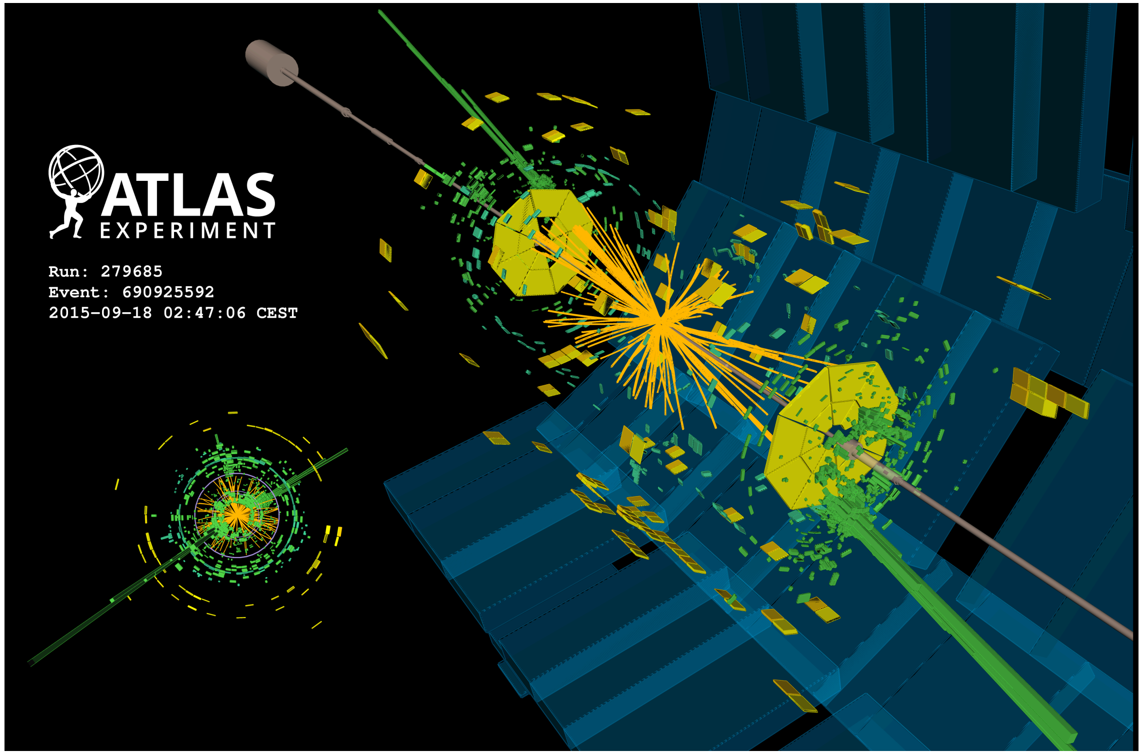
Typical LHC Workflow

Detector has O(100 million) sensors

Can't build 100M dimensional histogram

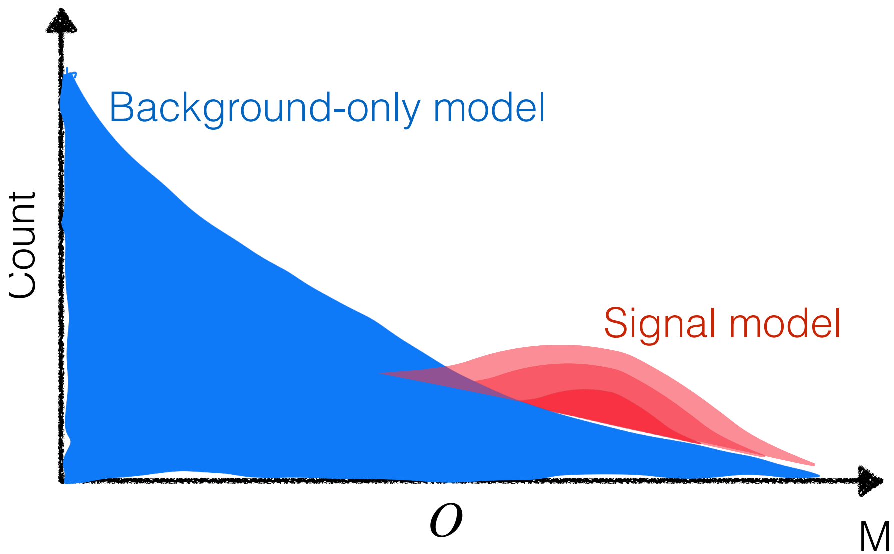
Reconstruction pipeline, event selection

Design sensitive one-dimensional observable

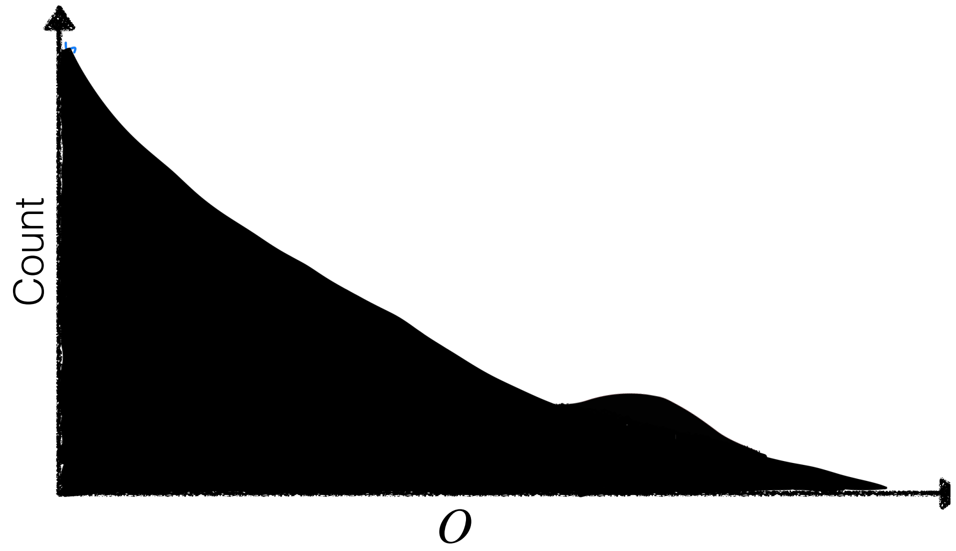


Density Estimation: What we’re used to doing..

Theory Predictions



Data

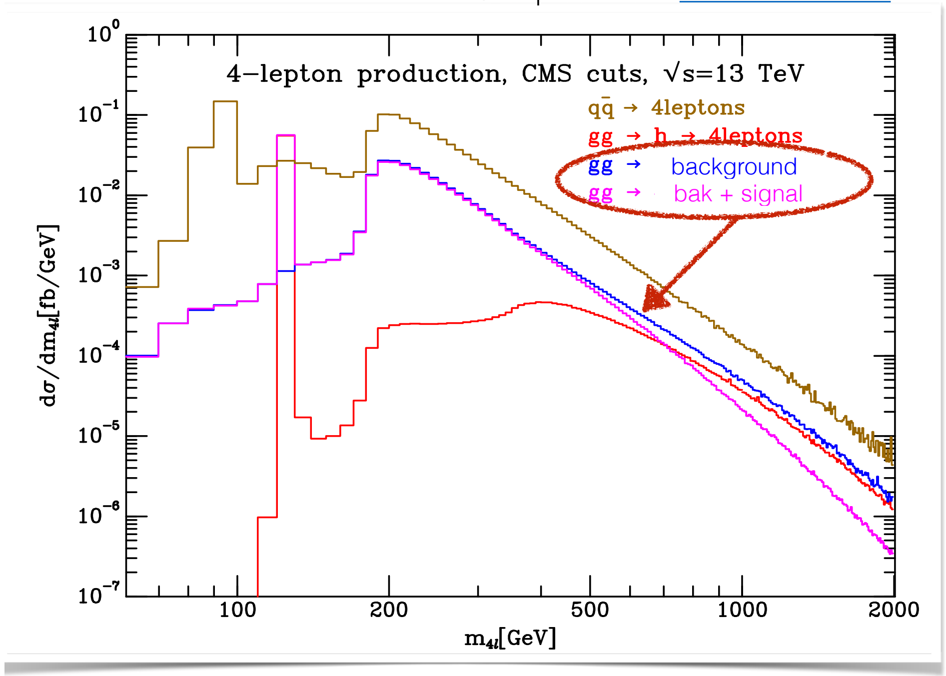


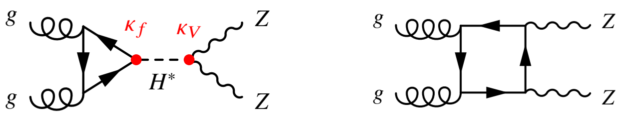
Measure signal strength μ

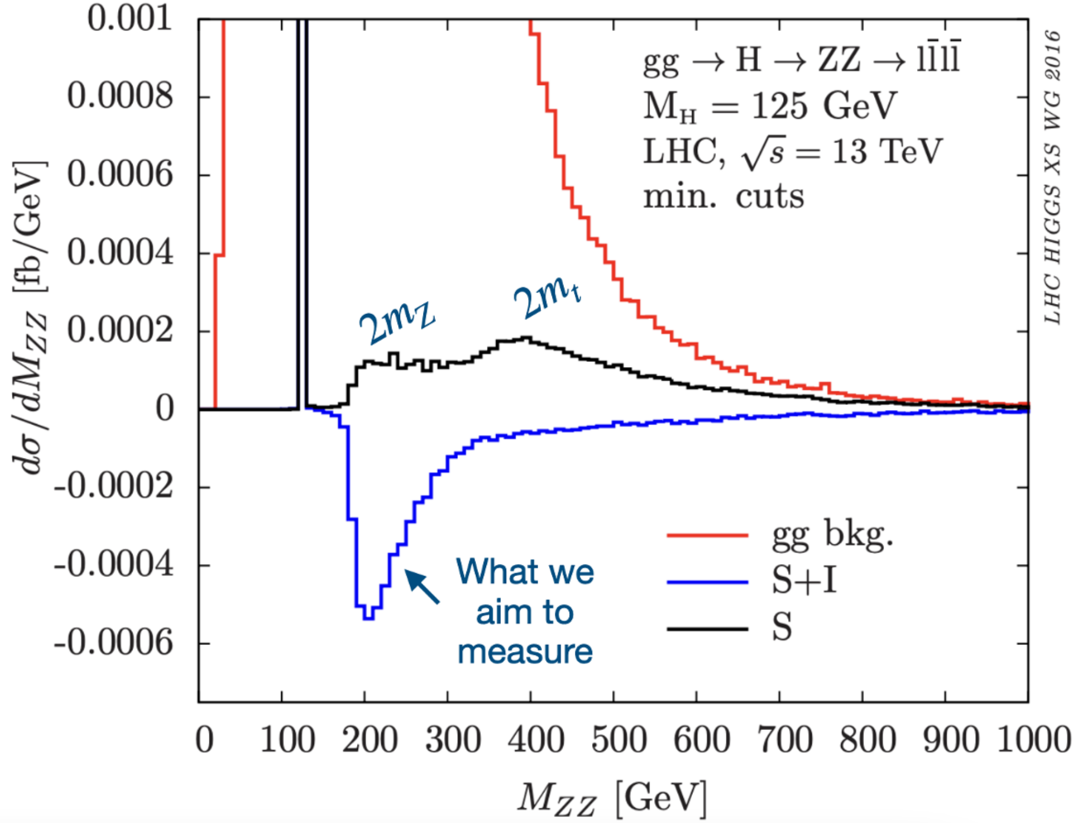
With histograms we can ask “Given the data, what is the likelihood of $\mu = 1$ hypothesis vs $\mu = 2$ hypothesis?”

New challenge: Non-linear changes in kinematics (w.r.t. parameter of interest)

Campbell et al: arXiv:1311.3589





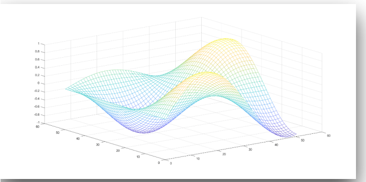


. histogram of any single observable is no longer optimal (see Ghosh et al: hal-02971995(p172)), but neural networks estimate high-dimensional likelihood ratios (see Cranmer et al: arXiv:1506.02169) !

Cranmer et al: [arXiv:1506.02169](https://arxiv.org/abs/1506.02169)

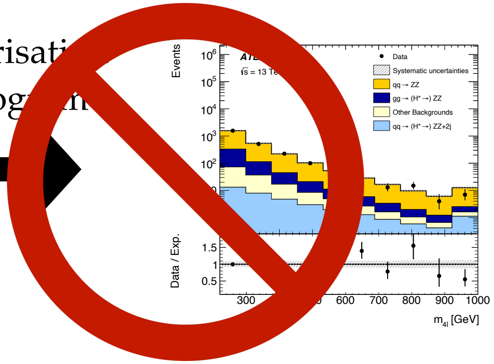
“Neural Simulation-Based Inference”

Traditional framework:



High-dim data

Summarisation to histogram



Summary Histogram →

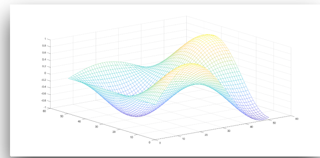
μ_1 →

Statistical Fit

→ Likelihood $\mathscr{L}(\mu_1 | \mathscr{D})$

μ is now arbitrary parameter of interest(s)

The neural inference framework:



High-dim data

Obs Data →

μ_1 →

Neural Networks

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

Open problems to extend to full ATLAS analysis:

- Robustness: Design and validation
- Systematic Uncertainties: Incorporate them in likelihood (ratio) model
- Neyman Construction: Sampling pseudo-experiments in a per-event analysis

Open problems to extend to full ATLAS analysis:

- Robustness: Design and validation
- Systematic Uncertainties: Incorporate them in likelihood (ratio) model
- Neyman Construction: Sampling pseudo-experiments in a per-event analysis

Search-Oriented Mixture Model

General Formula

Estimated using an ensemble of networks

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

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

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

Event rates estimated from simulations

Known analytically from theory model

Example use case

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

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

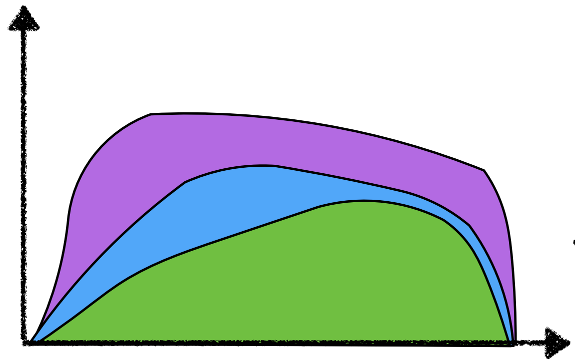
NN NN

Robust, parameterised classifier without parameterising

H_{ref} : Reference hypothesis

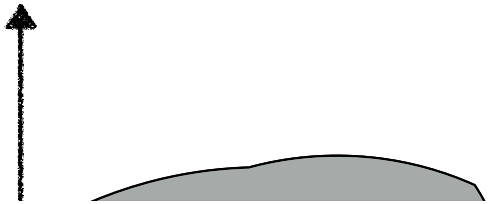
$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

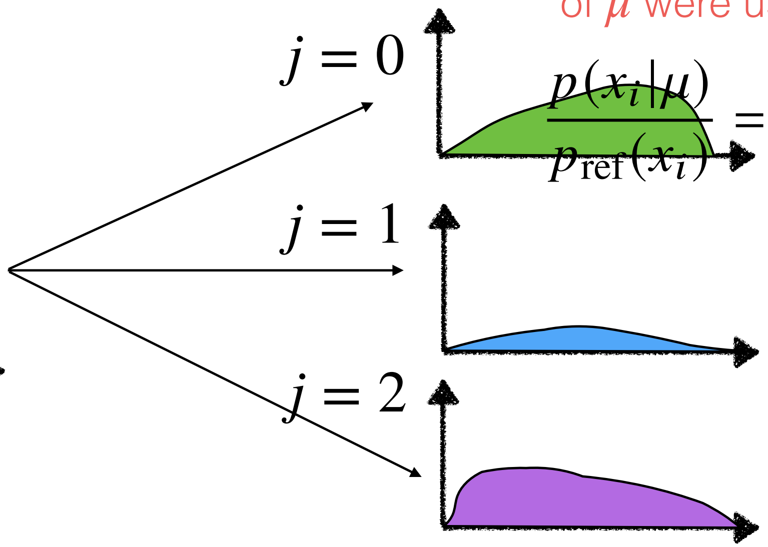
H_μ hypothesis



VS

H_{ref}



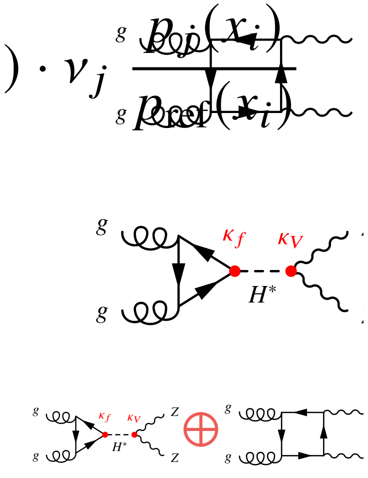


$$\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}$$

$f_0(\mu)$

$f_1(\mu) = \mu$

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



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

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

Reference Sample

A combination of signal samples, to ensure there's non-vanishing support in pre-selected region

$$p_{\text{ref}}(x_i) = \frac{1}{\sum_k \nu_k} \sum_k^{C_{\text{signals}}} \nu_k \cdot p_k(x_i)$$

⇒ In our dataset, $p_{ref}(\cdot) = p_S(\cdot)$

Choice of $p_{ref}(\cdot)$ can be made purely on numerical stability of training, as it drops out from the profile likelihood ratio

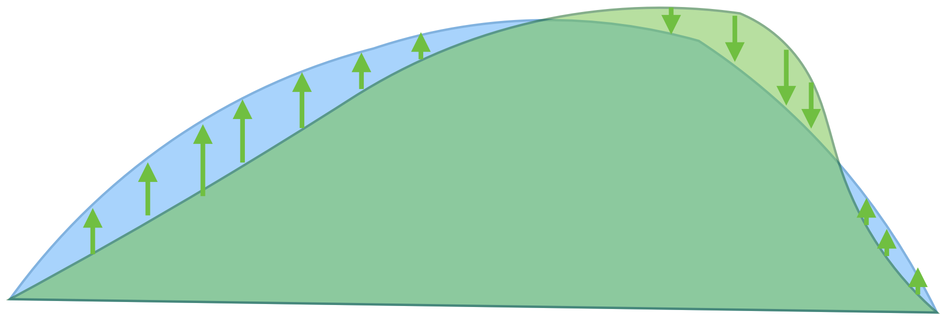
$$t_\mu = -2\ln\left(\frac{L_{\text{full}}(\mu, \widehat{\widehat{\alpha}})/\cancel{L_{\text{ref}}}}{L_{\text{full}}(\widehat{\mu}, \widehat{\alpha})/\cancel{L_{\text{ref}}}}\right)$$

Validate quality of LR estimation with re-weighting task

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

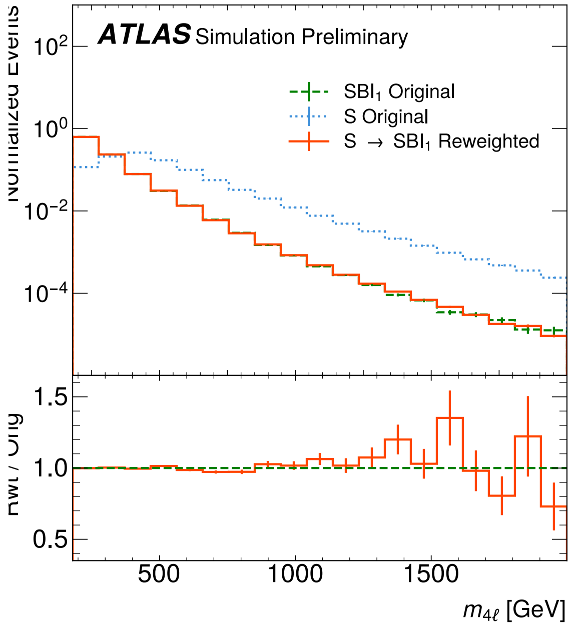
$$w_i = r_j(x_i) = \frac{p_j(x_i)}{p_{ref}(x_i)}$$

Already estimated using an ensemble of networks

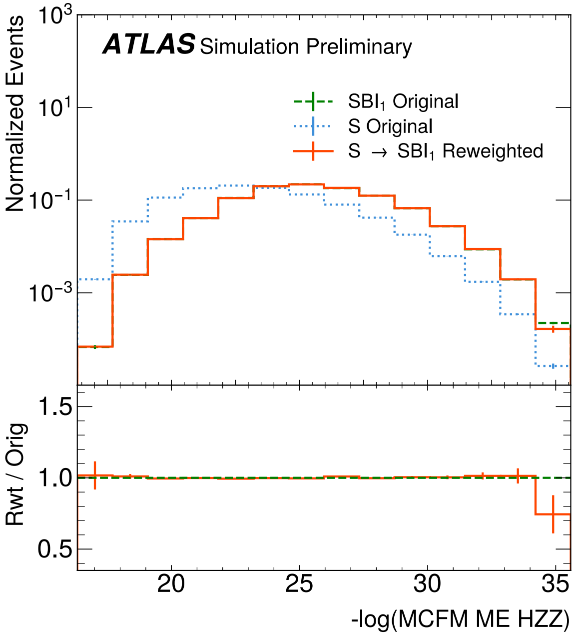


Re-weight closures

Source
Target
RW



m_{4l}



Matrix-Element-based Observable
(ggF from MCFM)

High-Dim Classifier Test:
Train independent classifier on RW vs Target, AUC=0.5 ⇒ LRs well estimated

ee additional closure tests

Open problems to extend to full ATLAS analysis:

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

Systematic uncertainties

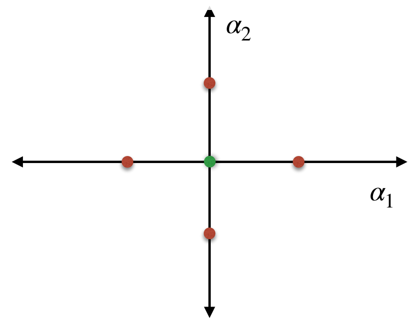
Experimental uncertainties:
Eg. Inaccuracies in the calibration of our detector

Image: arXiv:2105.08742

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

Image: arXiv:2109.08159

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



Known interpolation strategies

[See](#) formula used

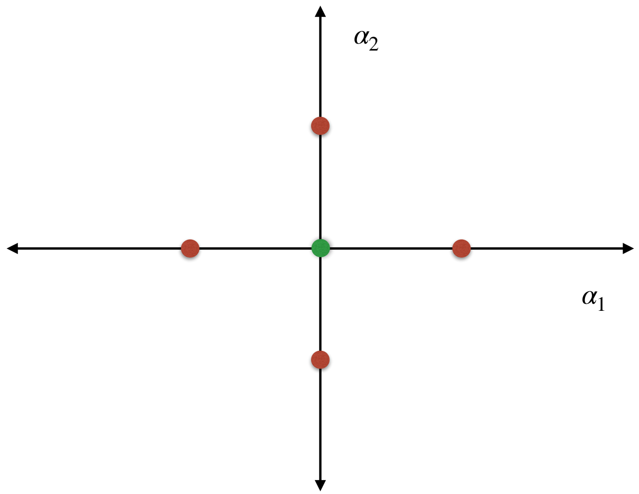
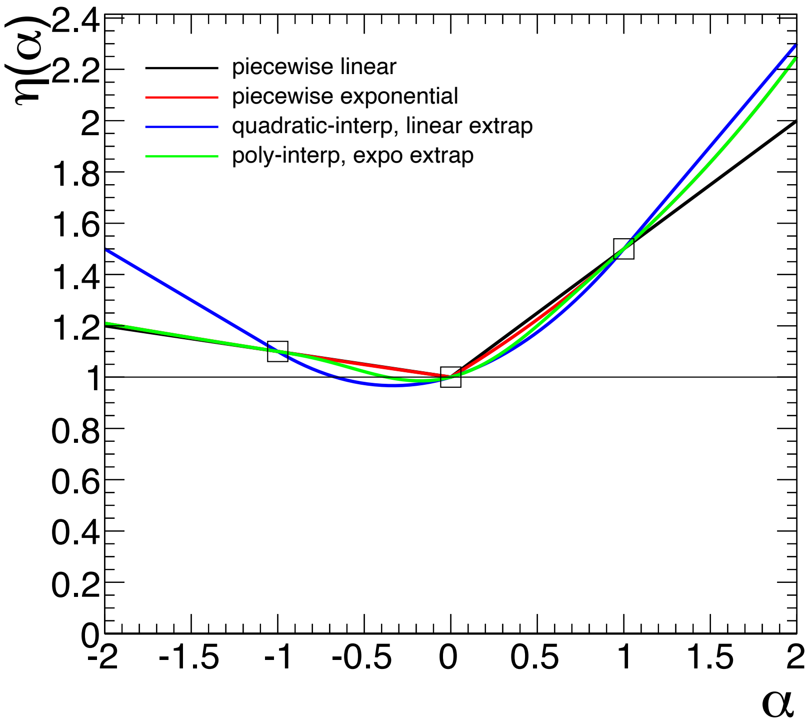


Image: arXiv:1503.07622



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

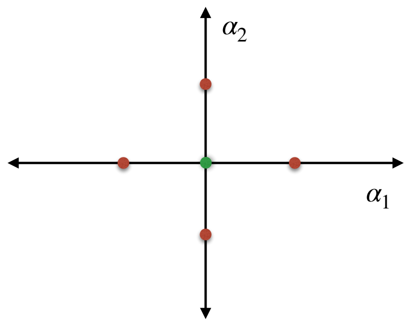
Probability density ratio including nuisance parameters (α)

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

We have this already

Estimated using another ensemble of networks and interpolation methods

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 $G_j(\alpha_k)$, $g_j(x_i,\alpha_k)$

Final test statistic

From previous slide

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

Rate term

Constrain term

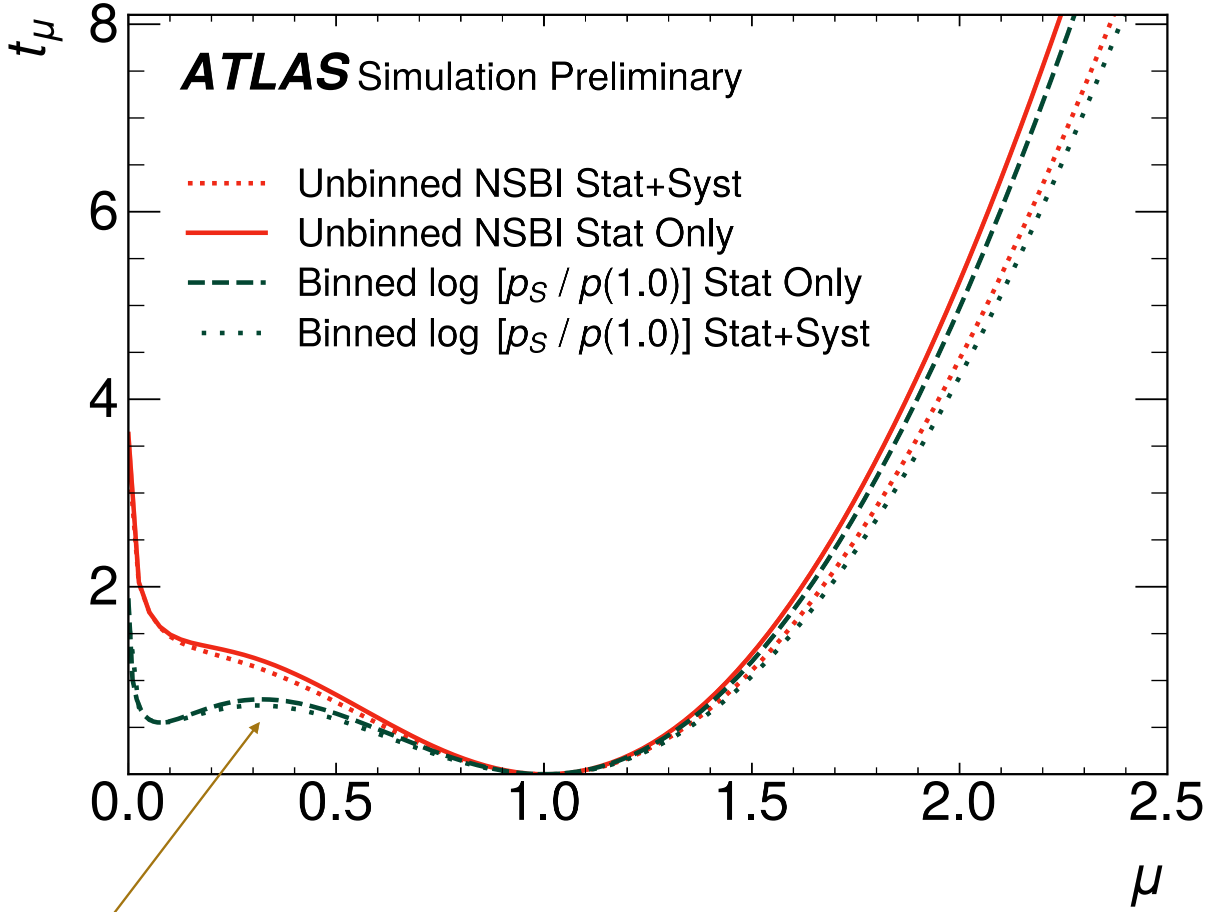
Prod over events

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

Negative Log Likelihood fit



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

Open problems to extend to full ATLAS analysis:

✓ Robustness: Design and validation

✓ Systematic Uncertainties: Incorporate them in likelihood (ratio) model

‣ Neyman Construction: Sampling pseudo-experiments in a per-event analysis

Asimov dataset

- Given ~10^6 MC events x_i= to simulate an experiment yielding ~1000 events=> normalize all events properly so that :

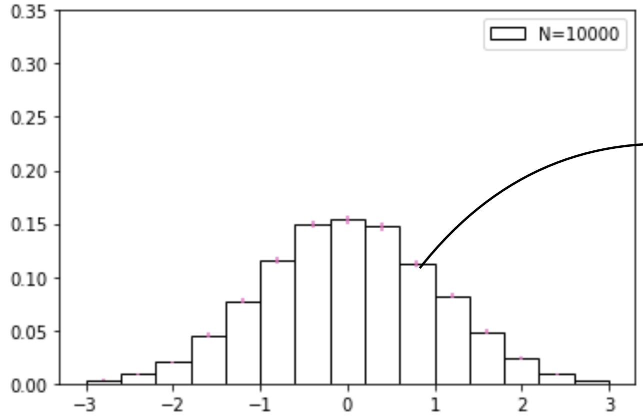
$$\sum_i w_i^{\text{Asimov}} \sim \mathcal{L}\sigma = \nu$$

- More generally,

$$\forall V \subset E, \sum_{x_i \in V} w_i^{\text{Asimov}} \sim \int_V \mathcal{L}\sigma(x)dx \sim \int_V \nu p(x)dx$$

Sampling (per-event) pseudo-experiments

Traditionally:

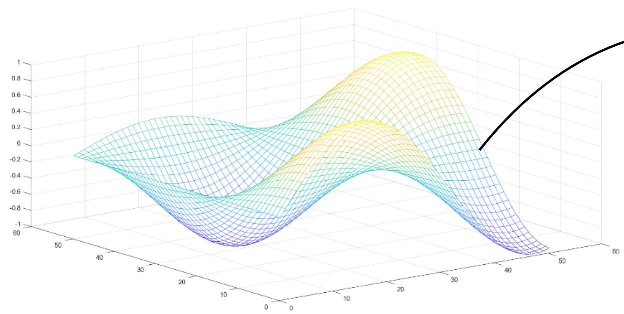


Asimov Histogram

Poisson per bin

$$N_i^{p.e} = Poisson(N_i^{Asimov})$$

NSBI:



Poisson per even

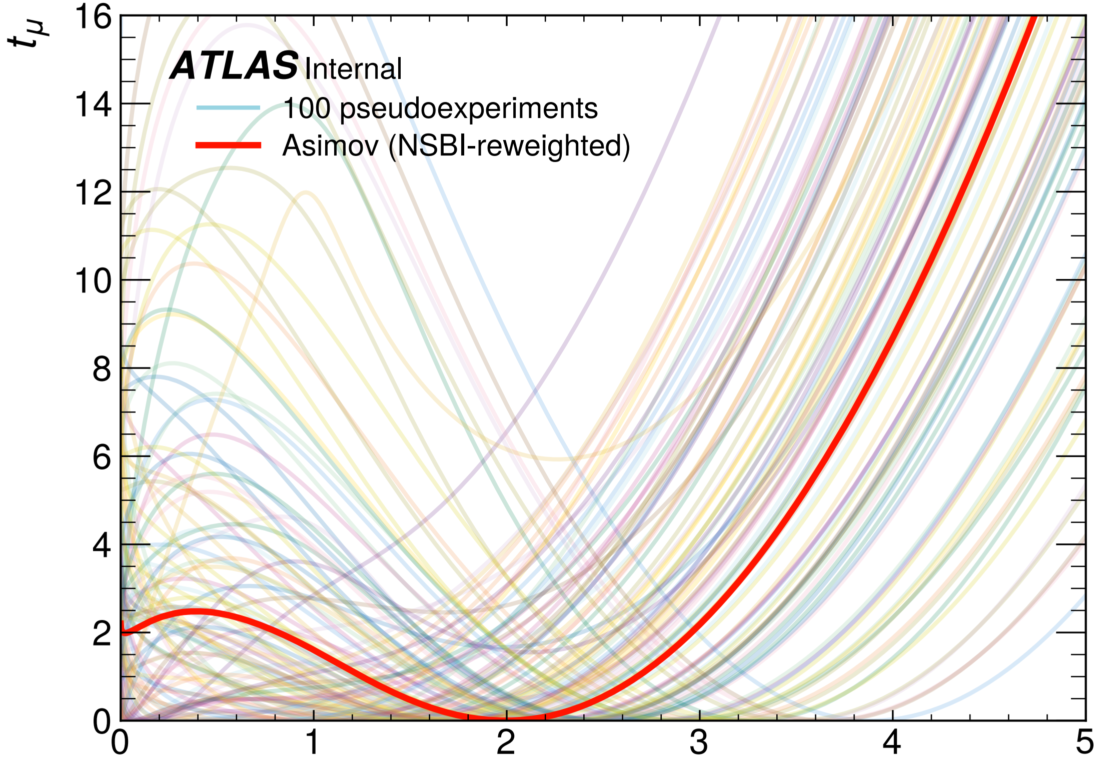
$$w_i^{p.e.} = Poisson(w_i^{Asimov})$$

(Somehow equivalent to drawing with replacement (bootstrap))

In practice given $w^{Asimov} < 1E\text{-}3$: w^{pe}=0, 1, rarely 2

Neyman Construction

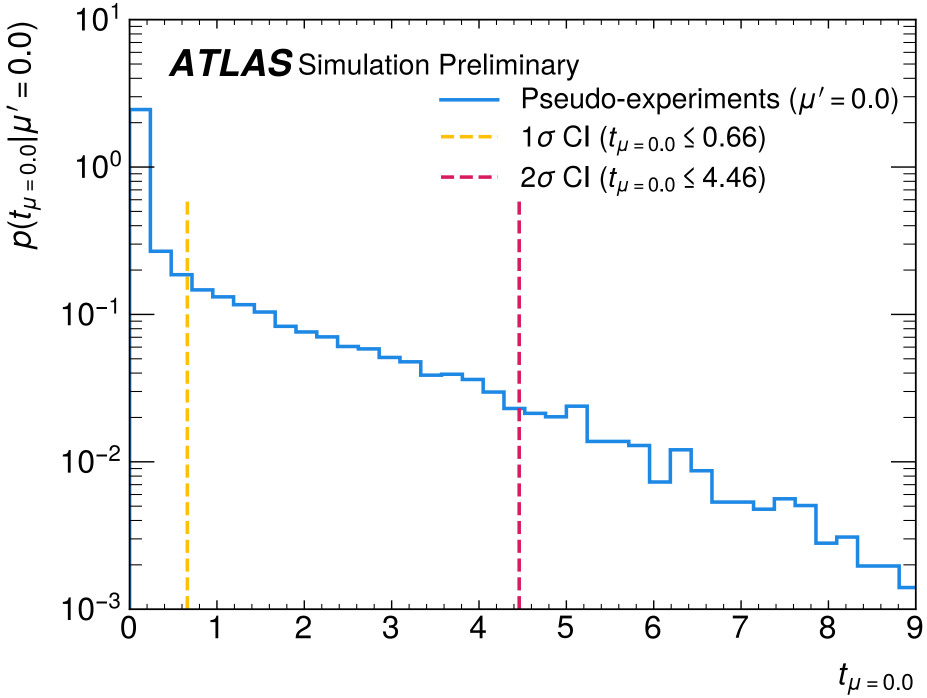
- Generate pseudo-experiments : identical to expected data taking, except for statistical fluctuations
- Fit Likelihood Ratio



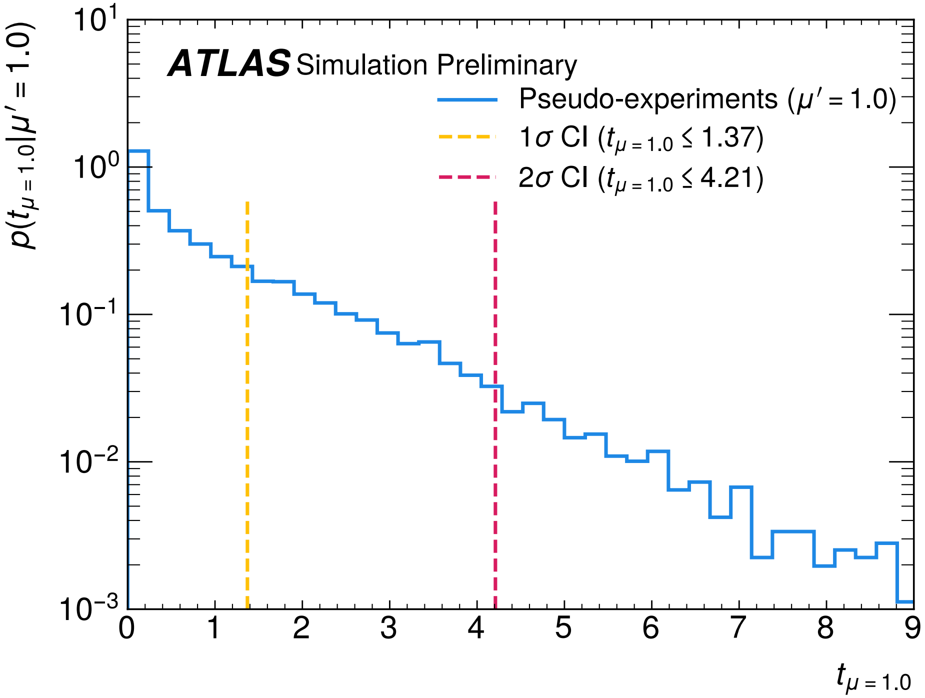
Neyman Construction

- To build confidence intervals, we need to 'invert the hypothesis test'
- Generate pseudo-experiments
- Compute the test statistic for the value of the hypothesis
- Integrate up to 68.27% (95.45%) to determine 1σ & 2σ CI as a function of parameter of interest

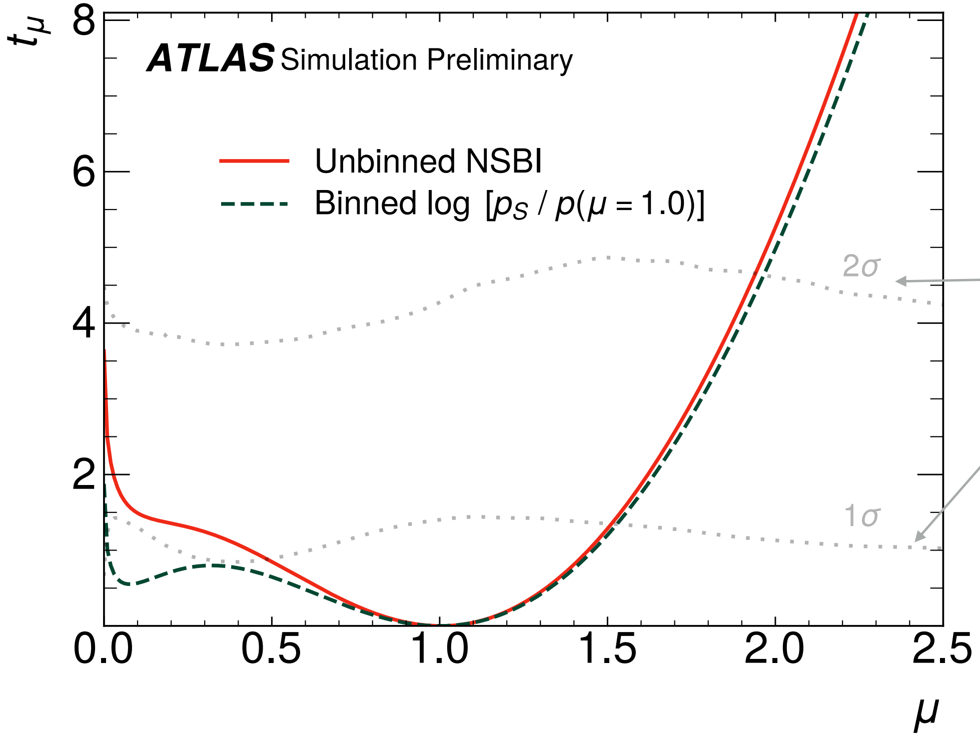
True $\mu = 0$



True $\mu = 1$



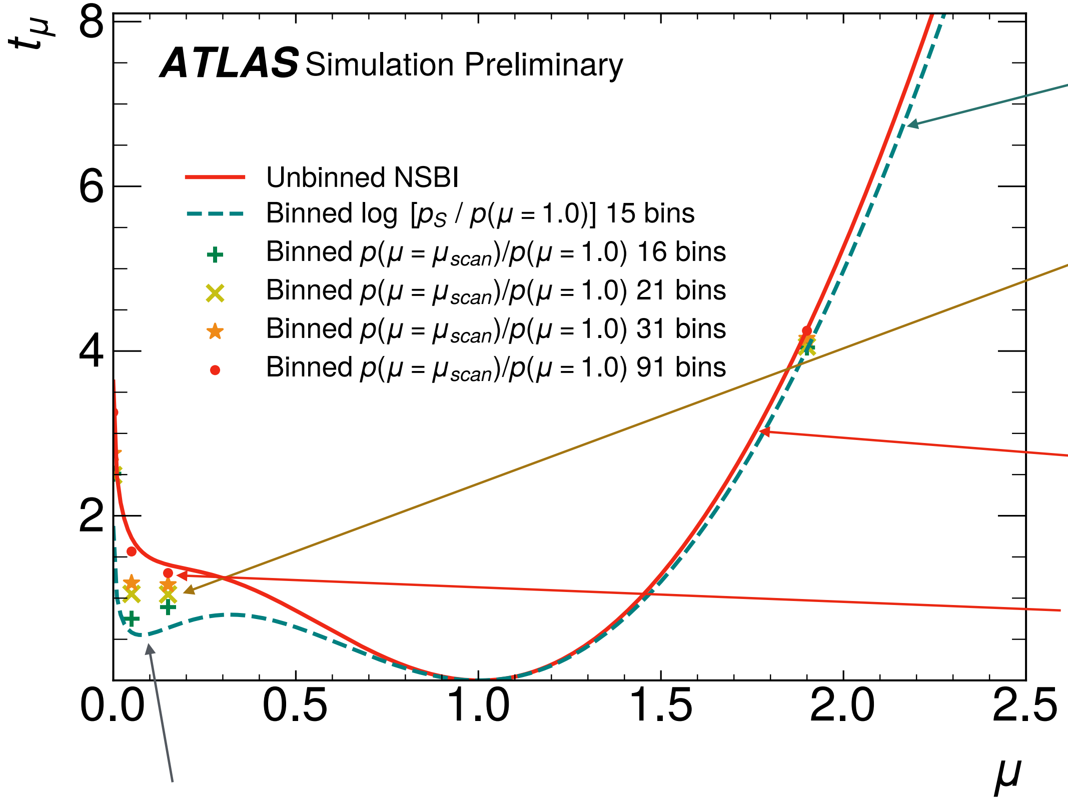
Confidence belts



Similar to structure seen in histogram analysis

Why does NSBI work better than traditional analyses?

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

$O_\mu = \frac{p(x_i \,|\, \mu)}{p(x_i \,|\, \mu = 1)}$: Parameterised observable, histogram fi

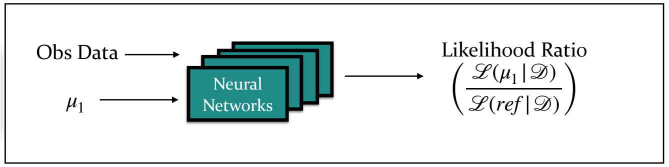
NSBI: Parameterised, unbinned

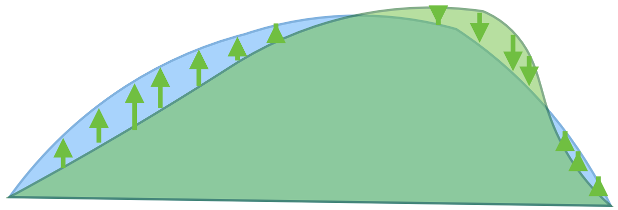
O_μ approaches NSBI as nBins $\rightarrow \infty$

Significant improvement in quantum interference impacted region

Conclusion

- Developed a complete statistical framework for high-dimensional statistical inference
    - Builds upon traditional methodology in ATLAS
    - Developed diagnostic tools for validation
- Such methods are crucial for analyses where kinematic distributions change non-linearly with the parameter of interest, eg. EFT studies
- Weaknesses: Similarly to traditional analyses, this method requires well trained neural networks and good Monte-Carlo simulation samples





Thanks!