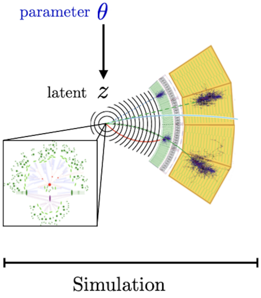
Advancing Tools for Simulation-Based Inference

Giovanni De Crescenzo
7 November 2024
[2410.07315]: Bahl, Bresó, De Crescenzo, Plehn

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

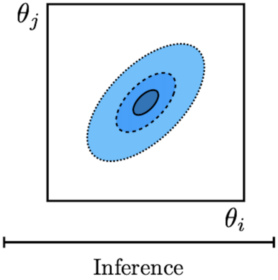
Simulation-Based Inference: basics

- Inference on θ theory parameters
- Intractable likelihood of the process $p(x|\theta)$
- Leverage θ dependent simulations: $r(x|\theta, \underbrace{\theta_0}_{\text{fixed}}) := \frac{p(x|\theta)}{p(x|\theta_0)}$



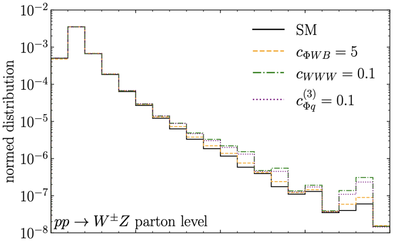
Estimate likelihood ratio $r(x|\theta,\ \theta_0)$

Sample test points to build Confidence Intervals

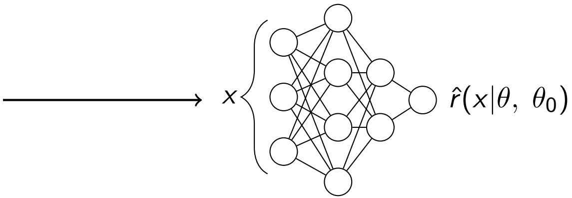


Standard approach and Machine Learning

- x-events are high dimensional $\mathcal{O}(10-100)$
- Standard approach relies on histogram/binned 1D analysis
- ML allows for *full output* analysis



Histogram binned likelihoods



Continuous approximation

Likelihood particle physics structure

- Events likelihood:

$$p(x|\theta) = \int d\xi dz_p \, p(x|\xi)p(\xi|z_p)p(z_p|\theta)$$

- ξ is the latent, high dim variable accounting for showering/hadronization/detector
- The integral makes the likelihood, thus the ratio intractable
- Notice factorization in the likelihood
- The joint ratio $r(x, z_p|\theta,\ \theta_0) = r(z_p|\theta,\ \theta_0)$ is then computable

Loss function

- Simple MSE loss, leveraging the simulated data

$$F[f_\varphi] = \int dx \int d\theta \int dz_p \underbrace{q(\theta)}_{\text{prior}} \underbrace{p(x|z_p)p(z_p|\theta)}_{\text{simulation likelihood}} \Bigg[\underbrace{f(z_p|\theta)}_{\text{labels}} - \underbrace{f_\varphi(x|\theta)}_{\text{NN}} \Bigg]^2$$

- Setting the *parton level labels* $f(z_p, \theta)$ defines the minimum for $f_\varphi(x, \theta)$
- e.g. $f(z_p, \theta) \equiv r(z_p|\theta_0, \theta) = \frac{1}{r(z_p|\theta,\theta_0)} = \frac{p(z_p|\theta_0)}{p(z_p|\theta)} \implies f_\varphi(x, \theta) \to r(x|\theta,\ \theta_0)$

Standard Model Effective Field Theory

- We consider now the SMEFT setup

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \sum_i \frac{c_i}{\Lambda^2}\, O_i \equiv \mathscr{L}_{\text{SM}} + \sum_i \theta_i\, O_i$$

- We consider an expansion in $1/\Lambda^2$ up to second order, making the differential cross section a *polynomial* of second degree

$$d\sigma(z_p|\theta) = d\sigma(z_p|\theta_0 = 0) + \sum_i \theta_i f_i(z) + \sum_{ij} \theta_i\theta_j f_{ij}(z_p)$$

Inductive bias

- Main idea: enforce the θ dependence in the learned quantity $r(x|\theta,\ \theta_0)$

We can enforce the θ dependence in two ways

- Derivative learning [2107.10859]

$$r(x|\theta,\ \theta_0) = 1 + \sum_i \theta_i \underbrace{R_i(x)}_{learnable} + \sum_{ij} \theta_i\theta_j \overbrace{R_{ij}(x)}^{learnable}$$

- Morphing aware [1805.00013]

$$r(x|\theta,\theta_0) = \vec{v}(\theta)M^{-1}(\theta)\vec{r}(x) \text{ with } r_i(x) = \underbrace{r(x|\theta_i,\theta_0)}_{learnable}$$

- This means dropping θ from $f_\varphi(x,\theta)$ and having separate $f_{\varphi,i}(x)$

Technical details

Sampling

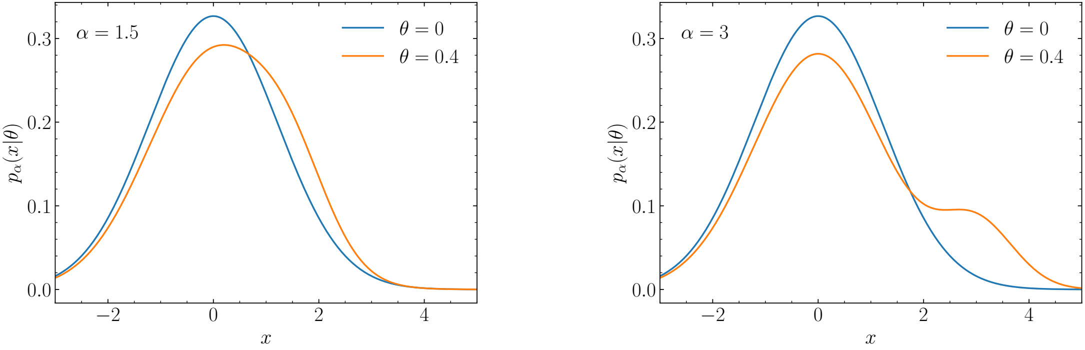
- In the derivative learning approach we sample only from the Standard Model point
- The morphing aware approach requires for each learnable function a composite sampling, half SM half from θ_i...
- ...thus populating more phase space

Problem: outliers

- Some events show large label values making training unstable
- Employed solution: fractional smearing (next slides)

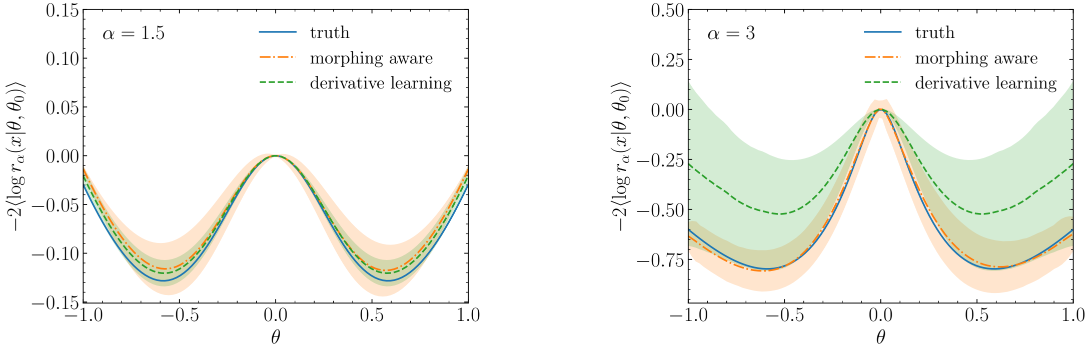
Toy model example: derivative learning and morphing aware

- Polynomial θ dependence (1D space)
- Known analytical likelihood: $p_\alpha(x|\theta) = \frac{\mathcal{N}_{0,1.22}(x)+\theta^2\mathcal{N}_{\alpha,0.71}(x)}{1+\theta^2}$



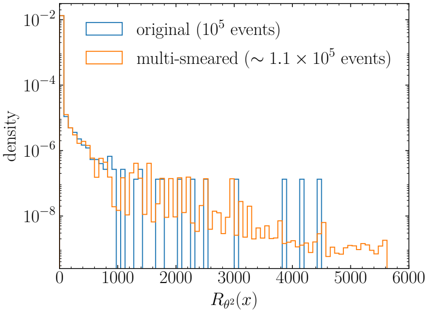
Toy model example: derivative learning and morphing aware

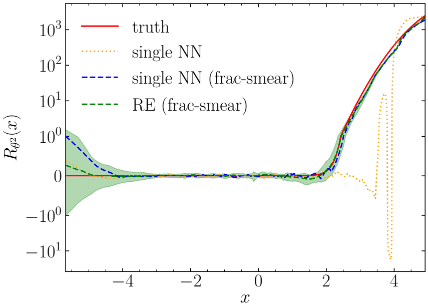
- Polynomial θ dependence (1D space)
- Known analytical likelihood: $p_\alpha(x|\theta) = \frac{\mathcal{N}_{0,1.22}(x)+\theta^2\mathcal{N}_{\alpha,0.71}(x)}{1+\theta^2}$
- Repulsive Ensembles for uncertainty quantification [2106.11642]



Toy model example: fractional smearing

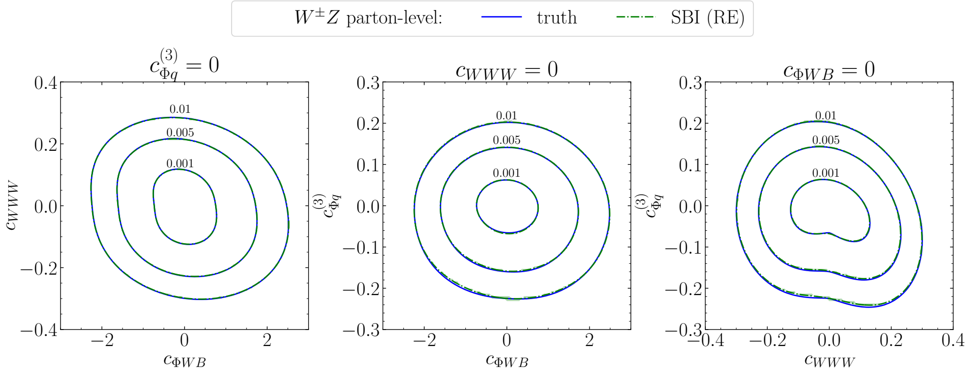
- Solve presence of outliers in the training
- Procedure:
  - Set label threshold for defining an outlier event z_p
  - Reweight z_p events $\frac{1}{n}$
  - Apply $p(x|z_p)$ smearing n times





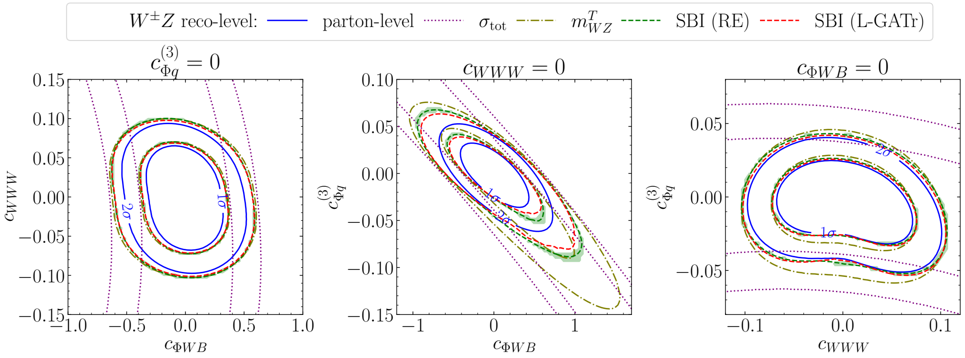
Physics results: parton level

- Consider 3 Wilson coefficients. 3D parameter space
- WZ production at $\sqrt{s} = 13.6$ TeV
- 300k training events
- Derivative learning approach only



Physics results: reconstruction level

- Define confidence intervals (contours)
- 650k training events
- Another application for L-GATr (see Jonas/Victor talks!)



Conclusions

- We can enforce the SMEFT structure in the way we learn likelihood ratios
  - Derivative Learning
  - Morphing aware
- Fractional smearing has proven to be a fundamental preprocessing step
- The particular parametrization for the networks can be realized by a simple MLP or the more task specific L-GATr architecture
- Set of tools for SMEFT analyses in LHC context

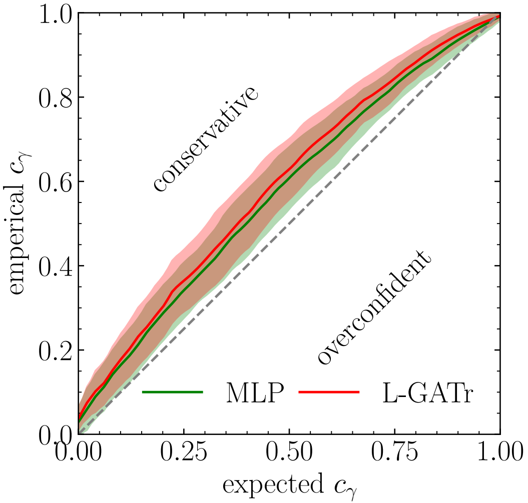
Backup: labels definition

- In the derivative learning approach the labels are defined as follows:

$$f(z_p|\theta) \equiv f(z_p) = R_i(z_p) \equiv \frac{\partial}{\partial\theta_i} \frac{d\sigma(z_p|\theta)/dz_p}{d\sigma(z_p|\theta_0)/dz_p}\bigg|_{\theta=\theta_0} = \frac{\partial_{\theta_i}|\mathcal{M}(z_p|\theta)|^2}{|\mathcal{M}(z_p|\theta_0)|^2}\bigg|_{\theta_0}$$

$$f(z_p|\theta) \equiv f(z_p) = R_{ij}(z_p) \equiv \frac{\partial^2}{\partial\theta_i\partial\theta_j} \frac{d\sigma(z_p|\theta)/dz_p}{d\sigma(z_p|\theta_0)/dz_p}\bigg|_{\theta=\theta_0} = \frac{\partial_{\theta_i}\partial_{\theta_j}|\mathcal{M}(z_p|\theta)|^2}{|\mathcal{M}(z_p|\theta)|^2}\bigg|_{\theta_0}$$

Backup: coverage



Backup: BSM testing point