Learning New Physics from an (Imperfect) Machine

Andrea Wulzer



Università degli Studi di Padova



Based on: D'Agnolo, AW, 2018 D'Agnolo, Grosso, Pierini, AW, Zanetti, 2019 D'Agnolo, Grosso, Pierini, AW, Zanetti, 2021

What if* the **RIGHT BSM model** has not been formulated? *very likely

What if* the **RIGHT BSM model** has not been formulated? *very likely

Would we still see the SM fail to describe data?

What if* the **RIGHT BSM model** has not been formulated? *very likely

Would we still see the SM fail to describe data?

Most likely not !

BSM is tiny departure from SM, or large in tiny prob. region Affecting few (unknown) observables over ∞ many we can measure

What if* the **RIGHT BSM model** has not been formulated? *very likely

Would we still see the SM fail to describe data?

Most likely not !

BSM is tiny departure from SM, or large in tiny prob. region Affecting few (unknown) observables over ∞ many we can measure

Regular New Physics searches are Model Dependent Choose observables sensitive to one BSM model This observable in general not sensitive to another BSM model

What if* the **RIGHT BSM model** has not been formulated?

Would we still see the SM fail to describe data?

Most likely not !

BSM is tiny departure from SM, or large in tiny prob. region Affecting few (unknown) observables over ∞ many we can measure

Regular New Physics searches are Model Dependent Choose observables sensitive to one BSM model This observable in general not sensitive to another BSM model



What if* the **RIGHT BSM model** has not been formulated? *very likely

Would we still see the SM fail to describe data?

"Regular" Model-Independence: weaken hypothesis on BSM nature, e.g.

- Simplified Model (of, say, SUSY, or DM, or HVT, ...)
- Effective Field Theories
- Bump Hunt

Re

"Machine-Learner" Model-Independence: eliminate phenomenological modelling altogether

We must design **Model Independent** searches aimed at detecting "generic" data departures from SM

What if* the **RIGHT BSM model** has not been formulated? *very likely

Would we still see the SM fail to describe data?

Most likely not !

BSM is tiny departure from SM, or large in tiny prob. region Affecting few (unknown) observables over ∞ many we can measure

Regular New Physics searches are Model Dependent Choose observables sensitive to one BSM model This observable in general not sensitive to another BSM model

> We must design **Model Independent** searches aimed at detecting "generic" data departures from SM SM = "Reference Model", to be compared with data without reference to alternative physics model

What if* the **RIGHT BSM model** has not been formulated? *very likely

Would we still see the SM fail to describe data?

Most likely not !

BSM is tiny departure from SM, or large in tiny prob. region Affecting few (unknown) observables over ∞ many we can measure

Regular New Physics searches are Model Dependent Choose observables sensitive to one BSM model This observable in general not sensitive to another BSM model

> Conceptually, same problem as assessing quality of a fit to data. AKA, **GoF Problem**

SM = "Reference Model", to be compared with data without reference to alternative physics model

(Foundation of entire LHC statistical practice)

Data: $\mathcal{D} = \{x_i\}, i = 1, \dots, \mathcal{N}_{\mathcal{D}}$

I.i.d. measurements of, e.g., reconstructed particle momenta in a region of interest

(Foundation of entire LHC statistical practice)

Data: $\mathcal{D} = \{x_i\}, i = 1, \dots, \mathcal{N}_{\mathcal{D}}$ Reference Distribution: $n(x|\mathbf{R})$ Alternative Distribution: $n(x|\mathbf{w})$ depending on **parameters** (composite)

$$n(x) = N P(x)$$
$$N = \int dx n(x)$$

(Foundation of entire LHC statistical practice)

Data:
$$\mathcal{D} = \{x_i\}, i = 1, \dots, \mathcal{N}_{\mathcal{D}}$$

Reference Distribution: $n(x|\mathbf{R})$
Alternative Distribution: $n(x|\mathbf{w})$
depending on **parameters** (composite)

$$n(x) = N P(x)$$
$$N = \int dx n(x)$$

Test statistic:
$$t(\mathcal{D}) = 2 \operatorname{Max} \left\{ \log \left[\frac{e^{-N}}{N} \right] \right\}$$

$$\operatorname{fax}_{\mathbf{w}} \left\{ \log \left[\frac{e^{-N(\mathbf{w})}}{e^{-N(\mathbf{R})}} \prod_{i=1}^{\mathcal{N}_{\mathcal{D}}} \frac{n(x_i | \mathbf{w})}{n(x_i | \mathbf{R})} \right] \right\}$$

Model Dependent Strategy

 $n(x|\mathbf{w}) = n(x|\mathrm{NP})$

Alternative as predicted by "NP" model. Few, or no, free parameters

(Foundation of entire LHC statistical practice)

Data:
$$\mathcal{D} = \{x_i\}, i = 1, \dots, \mathcal{N}_{\mathcal{D}}$$

Reference Distribution: $n(x|\mathbf{R})$
Alternative Distribution: $n(x|\mathbf{w})$
depending on **parameters** (composite)
Test statistic:

$$n(x) = N P(x)$$
$$N = \int dx n(x)$$

stic:

$$t(\mathcal{D}) = 2 \operatorname{Max}_{\mathbf{w}} \left\{ \log \left[\frac{e^{-N(\mathbf{w})}}{e^{-N(\mathbf{R})}} \prod_{i=1}^{\mathcal{N}_{\mathcal{D}}} \frac{n(x_i | \mathbf{w})}{n(x_i | \mathbf{R})} \right] \right\}$$

Model Dependent Strategy

 $n(x|\mathbf{w}) = n(x|\mathrm{NP})$

Alternative as predicted by "NP" model. Few, or no, free parameters

Model Independent Strategy

 $n(x|\mathbf{w}) = n(x|\mathbf{R}) e^{f(x;\mathbf{w})}$

Alternative in parametrised form. $f(x;\mathbf{w})$ is flexible function approximant

(Foundation of entire LHC statistical practice)

Data:
$$\mathcal{D} = \{x_i\}, i = 1, \dots, \mathcal{N}_{\mathcal{D}}$$

Reference Distribution: $n(x|\mathbf{R})$
Alternative Distribution: $n(x|\mathbf{w})$
depending on **parameters** (composite)

$$n(x) = N P(x)$$
$$N = \int dx n(x)$$

tic:

$$t(\mathcal{D}) = 2 \operatorname{Max}_{\mathbf{w}} \left\{ \log \left[\frac{e^{-N(\mathbf{w})}}{e^{-N(\mathbf{R})}} \prod_{i=1}^{\mathcal{N}_{\mathcal{D}}} \frac{n(x_i | \mathbf{w})}{n(x_i | \mathbf{R})} \right] \right\}$$

Model Dependent Strategy

 $n(x|\mathbf{w}) = n(x|\mathrm{NP})$

Alternative as predicted by "NP" model. Few, or no, free parameters

Model Independent Strategy

 $n(x|\mathbf{w}) = n(x|\mathbf{R}) e^{f(x;\mathbf{w})}$

Alternative in parametrised form. $f(x; \mathbf{w})$ is flexible function approximant

If $f(x; \mathbf{w})$ is piece-wise constant



(Foundation of entire LHC statistical practice)

Data:
$$\mathcal{D} = \{x_i\}, i = 1, ..., \mathcal{N}_{\mathcal{D}}$$

Reference Distribution: $n(x|\mathbf{R})$
Alternative Distribution: $n(x|\mathbf{w})$
depending on **parameters** (composite)

$$n(x) = N P(x)$$
$$N = \int dx n(x)$$

tic:

$$t(\mathcal{D}) = 2 \operatorname{Max}_{\mathbf{w}} \left\{ \log \left[\frac{e^{-N(\mathbf{w})}}{e^{-N(\mathbf{R})}} \prod_{i=1}^{\mathcal{N}_{\mathcal{D}}} \frac{n(x_i | \mathbf{w})}{n(x_i | \mathbf{R})} \right] \right\}$$

Model Dependent Strategy

 $n(x|\mathbf{w}) = n(x|\mathrm{NP})$

Alternative as predicted by "NP" model. Few, or no, free parameters

Model Independent Strategy

 $n(x|\mathbf{w}) = n(x|\mathbf{R}) e^{f(x;\mathbf{w})}$

Alternative in parametrised form. $f(x; \mathbf{w})$ is flexible function approximant

If $f(x; \mathbf{w})$ is a neural network



(Foundation of entire LHC statistical practice)

N P(x)dx n(x)

Data: $\mathcal{D} = \{x_i\}, i = 1, \dots, \mathcal{N}_{\mathcal{D}}$

Basic idea: $f(x; \mathbf{w}) = NN$

replace histograms with NN, literally!

Highly motivated attempt:

- NN "effective" flexible but smooth function approx.
- Often "sold" as alternative to hist. to fit distributions
- Better dimensionality scaling

Model Dependent Strategy

 $n(x|\mathbf{w}) = n(x|\mathrm{NP})$

Alternative as predicted by "NP" model Few, or no, free parameters

Model Independent Strategy

 $n(x|\mathbf{w}) = n(x|\mathbf{R}) e^{f(x;\mathbf{w})}$

Alternative in parametrised form. $f(x; \mathbf{w})$ is flexible function approximant

If $f(x; \mathbf{w})$ is a neural network



Maximum Likelihood Loss

Turn the evaluation of "t" into supervised training problem:

$$n(x|\mathbf{w}) = n(x|\mathbf{R}) e^{f(x;\mathbf{w})}$$

$$t(\mathcal{D}) = 2 \operatorname{Max}_{\mathbf{w}} \left\{ \log \left[\frac{e^{-N(\mathbf{w})}}{e^{-N(\mathbf{R})}} \prod_{i=1}^{\mathcal{N}_{\mathcal{D}}} \frac{n(x_i|\mathbf{w})}{n(x_i|\mathbf{R})} \right] \right\} \stackrel{\bullet}{=} -2 \operatorname{Min}_{\mathbf{w}} \left[N(\mathbf{w}) - N(\mathbf{R}) - \sum_{i=1}^{\mathcal{N}_{\mathcal{D}}} f(x_i;\mathbf{w}) \right]$$
We need a **Reference Sample**, distributed according to Reference Model

$$\mathcal{R} = \{x_i\}, \ i = 1, \dots, \mathcal{N}_{\mathcal{R}}$$

Approximate integral as Monte Carlo sum:

$$N(\mathbf{w}) = \int dx \, n(x|\mathbf{R}) \, e^{f(x;\mathbf{w})} = \frac{N(\mathbf{R})}{\mathcal{N}_{\mathcal{R}}} \sum_{x \in \mathcal{R}} e^{f(x;\mathbf{w})}$$

Maximum Likelihood Loss

Turn the evaluation of "t" into supervised training problem:

$$n(x|\mathbf{w}) = n(x|\mathbf{R}) e^{f(x;\mathbf{w})}$$

$$t(\mathcal{D}) = 2 \operatorname{Max}_{\mathbf{w}} \left\{ \log \left[\frac{e^{-N(\mathbf{w})}}{e^{-N(\mathbf{R})}} \prod_{i=1}^{\mathcal{N}_{\mathcal{D}}} \frac{n(x_i|\mathbf{w})}{n(x_i|\mathbf{R})} \right] \right\} \stackrel{\bullet}{=} -2 \operatorname{Min}_{\mathbf{w}} \left[N(\mathbf{w}) - N(\mathbf{R}) - \sum_{i=1}^{\mathcal{N}_{\mathcal{D}}} f(x_i;\mathbf{w}) \right]$$
We need a **Reference Sample**, distributed according to Reference Model

$$\mathcal{R} = \{x_i\}, \ i = 1, \dots, \mathcal{N}_{\mathcal{R}}$$

Approximate integral as Monte Carlo sum:

$$N(\mathbf{w}) = \int dx \, n(x|\mathbf{R}) \, e^{f(x;\mathbf{w})} = \frac{N(\mathbf{R})}{\mathcal{N}_{\mathcal{R}}} \sum_{x \in \mathcal{R}} e^{f(x;\mathbf{w})}$$

In order to read this as "equal", we need

$$\mathcal{N}_{\mathcal{R}} \gg N(\mathbf{R})$$

Like saying that $n(x | \mathbf{R})$ is "known", as it is infinitely samplable

$$\int_{X} \frac{1}{\sqrt{2}} \int_{X} \frac{1}{\sqrt{2}} \int_{X$$

$$\mathcal{N}_{\mathcal{R}} \stackrel{\sim}{\underset{x \in \mathcal{R}}{\overset{\sim}}} \mathcal{N}_{\mathcal{R}}$$

Get t = -2 * minimal loss. Trained net is fit to distribution log ratio

$$t(\mathcal{D}) = -2 \operatorname{Min}_{\{\mathbf{w}\}} \left[\frac{N(\mathbf{R})}{\mathcal{N}_{\mathcal{R}}} \sum_{x \in \mathcal{R}} (e^{f(x;\mathbf{w})} - 1) - \sum_{x \in \mathcal{D}} f(x;\mathbf{w}) \right] \equiv -2 \operatorname{Min}_{\{\mathbf{w}\}} L[f(\cdot,\mathbf{w})]$$
$$L[f] = \sum_{(x,y)} \left[(1-y) \frac{N(\mathbf{R})}{\mathcal{N}_{\mathcal{R}}} (e^{f(x)} - 1) - y f(x) \right]$$

The Algorithm

We compute "t" by supervised training using "ML-Loss"

Observed (or Toy) Data are class "1"

• Class "0" is a Reference Sample

SM-distributed synthetic instances of the features "x" Can come from **Monte Carlo**, or **Data Driven** Nothing different from "**background sample**" in regular searches

Preferably, more abundant than the data: $\mathcal{N}_{\mathcal{R}} \gg N(\mathbf{R})$

The Algorithm



The Algorithm

We compute "t" by supervised training using "ML-Loss"

Observed (or Toy) Data are class "1"

• Class "0" is the Reference Sample

SM-distributed synthetic instances of the features "x" Can come from **Monte Carlo**, or **Data Driven** Nothing different from "**background sample**" in regular searches Preferably, more abundant than the data: $\mathcal{N}_{\mathcal{R}} \gg N(\mathbf{R})$

We generate Toy Datasets in Reference Hypothesis, train on each and compute empirical P(t|R)

This will give us the observed p-value:

$$p = \int_{t_{\rm obs}} P(t|\mathbf{R})$$

Illustrating Performances



Illustrating Performances

(Simple 1d example with exponential Reference)



Distribution of "t" in one New Physics Model Hypothesis t \rightarrow p \rightarrow Z-score (we use $Z = \Phi^{-1}(1-p)$)





Reference Model Predictions are unavoidably imperfect e.g., PDF/Lumi/Detector Modeling ...

Imperfections are Nuisance Parameters

Constrained by **Auxiliary Measurements** Define a **composite** Reference hypothesis

Reference Model Predictions are unavoidably imperfect e.g., PDF/Lumi/Detector Modeling ...

Imperfections are Nuisance Parameters

Constrained by **Auxiliary Measurements** Define a **composite** Reference hypothesis

 $t(\mathcal{D}, \mathcal{A}) = 2 \log \frac{\max_{\mathbf{w}, \boldsymbol{\nu}} \left[\mathcal{L}(\mathbf{H}_{\mathbf{w}, \boldsymbol{\nu}} | \mathcal{D}) \cdot \mathcal{L}(\boldsymbol{\nu} | \mathcal{A}) \right]}{\max_{\boldsymbol{\nu}} \left[\mathcal{L}(\mathbf{R}_{\boldsymbol{\nu}} | \mathcal{D}) \cdot \mathcal{L}(\boldsymbol{\nu} | \mathcal{A}) \right]}$

 $H_{\mathbf{w},\boldsymbol{\nu}}$

Just like in no-nuisance case:

$$n(x|\mathbf{H}_{\mathbf{w},\boldsymbol{\nu}}) = e^{f(x;\mathbf{w})}n(x|\mathbf{R}_{\boldsymbol{\nu}})$$

Beyond-Reference effects parametrised by NN

Reference Model Predictions are unavoidably imperfect e.g., PDF/Lumi/Detector Modeling ...

Imperfections are Nuisance Parameters

Constrained by **Auxiliary Measurements** Define a **composite** Reference hypothesis

$$t(\mathcal{D}, \mathcal{A}) = 2 \max_{\mathbf{w}, \boldsymbol{\nu}} \log \left[\frac{\mathcal{L}(\mathbf{H}_{\mathbf{w}, \boldsymbol{\nu}} | \mathcal{D})}{\mathcal{L}(\mathbf{R}_{0} | \mathcal{D})} \cdot \frac{\mathcal{L}(\boldsymbol{\nu} | \mathcal{A})}{\mathcal{L}(\mathbf{0} | \mathcal{A})} \right] - 2 \max_{\boldsymbol{\nu}} \log \left[\frac{\mathcal{L}(\mathbf{R}_{\boldsymbol{\nu}} | \mathcal{D})}{\mathcal{L}(\mathbf{R}_{0} | \mathcal{D})} \cdot \frac{\mathcal{L}(\boldsymbol{\nu} | \mathcal{A})}{\mathcal{L}(\mathbf{0} | \mathcal{A})} \right]$$

$$H_{\mathbf{w}, \boldsymbol{\nu}}$$

$$t(\mathcal{D}, \mathcal{A}) = \tau(\mathcal{D}, \mathcal{A}) - \Delta(\mathcal{D}, \mathcal{A})$$
Central-Value Reference: R₀
Nuisance set to their C-V

"Tau" term by training on Data Almost like for no nuisance, but with modified ML-Loss:

$$L\left[f(\cdot;\mathbf{w}),\,\boldsymbol{\nu};\,\widehat{\delta}(\cdot)\right] = -\sum_{x_i\in\mathcal{D}}\left[f(x_i;\mathbf{w}) + \log(r(x_i;\boldsymbol{\nu}))\right] + \sum_{e\in\mathcal{R}} w_e\left[e^{f(x_e;\mathbf{w}) + \log(r(x_e;\boldsymbol{\nu}))} - 1\right] \\ + \log\left[\frac{\mathcal{L}(\boldsymbol{\nu}|\mathcal{A})}{\mathcal{L}(\mathbf{0}|\mathcal{A})}\right]$$

And, with simultaneous **training over the nuisance** parameters Data trained against **Central-Value Reference** sample **only**

$$t(\mathcal{D}, \mathcal{A}) = \tau(\mathcal{D}, \mathcal{A}) - \Delta(\mathcal{D}, \mathcal{A})$$

Reference Model Predictions are unavoidably imperfect e.g., PDF/Lumi/Detector Modeling ...

Imperfections are Nuisance Parameters

Constrained by **Auxiliary Measurements** Define a **composite** Reference hypothesis

$$t(\mathcal{D}, \mathcal{A}) = 2 \max_{\mathbf{w}, \boldsymbol{\nu}} \log \left[\frac{\mathcal{L}(\mathbf{H}_{\mathbf{w}, \boldsymbol{\nu}} | \mathcal{D})}{\mathcal{L}(\mathbf{R}_{0} | \mathcal{D})} \cdot \frac{\mathcal{L}(\boldsymbol{\nu} | \mathcal{A})}{\mathcal{L}(\mathbf{0} | \mathcal{A})} \right] - 2 \max_{\boldsymbol{\nu}} \log \left[\frac{\mathcal{L}(\mathbf{R}_{\boldsymbol{\nu}} | \mathcal{D})}{\mathcal{L}(\mathbf{R}_{0} | \mathcal{D})} \cdot \frac{\mathcal{L}(\boldsymbol{\nu} | \mathcal{A})}{\mathcal{L}(\mathbf{0} | \mathcal{A})} \right]$$

$$\mathbf{H}_{\mathbf{w}, \boldsymbol{\nu}}$$

$$\mathbf{H}_{\mathbf{w}, \boldsymbol{\nu}}$$

$$\mathbf{R}_{\boldsymbol{\nu}}$$

$$\mathbf{R}_{\boldsymbol{\nu}$$

$$\mathbf{R}_{\boldsymbol{\nu}}$$

$$\mathbf{R}_{\boldsymbol{\nu}}$$

Reference Model Predictions are unavoidably imperfect e.g., PDF/Lumi/Detector Modeling ...

Imperfections are Nuisance Parameters

Constrained by **Auxiliary Measurements** Define a **composite** Reference hypothesis

$$t(\mathcal{D}, \mathcal{A}) = 2 \max_{\mathbf{w}, \boldsymbol{\nu}} \log \left[\frac{\mathcal{L}(\mathbf{H}_{\mathbf{w}, \boldsymbol{\nu}} | \mathcal{D})}{\mathcal{L}(\mathbf{R}_{0} | \mathcal{D})} \cdot \frac{\mathcal{L}(\boldsymbol{\nu} | \mathcal{A})}{\mathcal{L}(\mathbf{0} | \mathcal{A})} \right] - 2 \max_{\boldsymbol{\nu}} \log \left[\frac{\mathcal{L}(\mathbf{R}_{\boldsymbol{\nu}} | \mathcal{D})}{\mathcal{L}(\mathbf{R}_{0} | \mathcal{D})} \cdot \frac{\mathcal{L}(\boldsymbol{\nu} | \mathcal{A})}{\mathcal{L}(\mathbf{0} | \mathcal{A})} \right]$$

$$H_{\mathbf{w}, \boldsymbol{\nu}}$$

$$H_{\mathbf{w}, \boldsymbol{\nu}}$$

$$\mathbf{R}_{\boldsymbol{\nu}}$$

$$t(\mathcal{D}, \mathcal{A}) = \tau(\mathcal{D}, \mathcal{A}) - \Delta(\mathcal{D}, \mathcal{A})$$
If we do all right, by Wilks-Wald we get:
$$P(t | \mathbf{R}_{\boldsymbol{\nu}}) = P(t | \mathbf{R}_{\mathbf{0}}) = \chi_{d}^{2}$$
Independence of t distribution on the true value of nuisance is essential for feasible test

An Imperfect Machine at Work

(Simple 1d example with exponential Reference)

Tau distribution distorted by non-central value nuisance if not corrected, produces false positives

t = Tau-Delta independent of nuisance

Remark #1: By Wilks-Wald Theorem, P(t|R) is a χ², with as many d.o.f. as fit parameters (for us, number of NN pars)... Provided statistics is large relative to "complexity" of model being fitted

or, which is the same

Provided fit model "simple enough", for given data stat.

We use **x²-compatibility** as **Model Selection criterion** Asy.For. violation = sensitivity to low-statistics portion of dataset = overfitting Selection w/o nuisance ensures nuisance-independent chi-sq Criterion used in particular to select **Weight Clipping** regularisation par.

Weight Clipping Selection

Remark #1: By Wilks-Wald Theorem, P(t|R) is a χ², with as many d.o.f. as fit parameters (for us, number of NN pars)... Provided statistics is large relative to "complexity" of model being fitted

or, which is the same

Provided fit model "simple enough", for given data stat.

We use **x²-compatibility** as **Model Selection criterion** Asy.For. violation = sensitivity to low-statistics portion of dataset = overfitting Selection w/o nuisance ensures nuisance-independent chi-sq Criterion used in particular to select **Weight Clipping** regularisation par.

Concern #1: We do not like Weight Clipping, and we would like better regularization and measure of NN complexity

Remark #2:

The Reference Sample is not of course infinite.

- We do empirically check that results weakly depend on the specific Reference sample instance.
- Factor few more abundant than Data found enough

Remark #2:

The Reference Sample is not of course infinite.

- We do empirically check that results weakly depend on the specific Reference sample instance.
- Factor few more abundant than Data found enough

Concern #2:

We have no Analytic/Asymptotic control of the Reference Sample fluctuations effects.

Remark #3:

Ours is a GoF 2-sample test from classifier training.

[see J.Friedman, 2004]

With specific test statistics and loss function choice, dictated by Maximum Likelihood approach.

Maximum Likelihood convenient viewpoint to deal with imperfections as nuisance parameters.

Remark #3:

Ours is a GoF 2-sample test from classifier training.

[see J.Friedman, 2004]

With specific test statistics and loss function choice, dictated by Maximum Likelihood approach.

Maximum Likelihood convenient viewpoint to deal with imperfections as nuisance parameters.

Concern #3:

No concern here.

But we should look for concrete GoF problems to try NPLM

Strategy has been defined, and applied to problems of the same scale of complexity as LHC analysis

Further progress requires full-fledged implementation in realistic LHC final state (2 leptons?, 4 leptons?, more exotic?)

- Strategy has been defined, and applied to problems of the same scale of complexity as LHC analysis
- Further progress requires full-fledged implementation in realistic LHC final state (2 leptons?, 4 leptons?, more exotic?)

Expected implementation challenges (limit on lumi. we can handle)

- Statistically accurate enough (large or smart) Reference Sample
- Generation of Reference-distributed Toys
- Accurate learning of nuisance Likelihood
- Training execution time

Strategy has been defined, and applied to problems of the same scale of complexity as LHC analysis

Further progress requires full-fledged implementation in realistic LHC final state (2 leptons?, 4 leptons?, more exotic?)

Expected implementation challenges (limit on lumi. we can handle)

- Statistically accurate enough (large or smart) Reference Sample
- Generation of Reference-distributed Toys
- Accurate learning of nuisance Likelihood
- Training execution time

Faster/Smarter Monte Carlo weighted samples generative models fast (but accurate) detector sim. Toys at NLO Generic need for the whole

HL-LHC analysis program!

Strategy has been defined, and applied to problems of the same scale of complexity as LHC analysis

Further progress requires full-fledged implementation in realistic LHC final state (2 leptons?, 4 leptons?, more exotic?)

Expected implementation challenges (limit on lumi. we can handle)

- Statistically accurate enough (large or smart) Reference Sample
- Generation of Reference-distributed Toys
- Accurate learning of nuisance Likelihood
- Training execution time

Faster/Smarter Monte Carlo weighted samples generative models fast (but accurate) detector sim. Toys at NLO Generic need for the whole

HL-LHC analysis program!

Likelihood-free Inference Techniques being worked out for EFT (MadMiner) **Stimulate** and **exploit** these developments

Strategy has been defined, and applied to problems of the same scale of complexity as LHC analysis

Further progress requires full-fledged implementation in realistic LHC final state (2 leptons?, 4 leptons?, more exotic?)

Expected implementation challenges (limit on lumi. we can handle)

- Statistically accurate enough (large or smart) Reference Sample.
- Generation of Reference-distributed Toys
 - Accurate learning of nuisance Likelihood

Non-NN Models Kernel Method "Falkon" [Letizia, Grosso, et. al., 2022]

Training execution time

Faster/Smarter Monte Carlo weighted samples generative models fast (but accurate) detector sim. Toys at NLO Generic need for the whole

HL-LHC analysis program!

Likelihood-free Inference Techniques being worked out for EFT (MadMiner) **Stimulate** and **exploit** these developments

Model-Independent search algorithms also good for:

- Comparison between Monte Carlo Generators
- Data Validation
- GoF

Model-Independent search algorithms also good for:

- Comparison between Monte Carlo Generators
- Data Validation
- GoF

When and if these techniques make it to real analyses, I suspect we will find plenty of wrong bck estimates ...

Model-Independent search algorithms also good for:

- Comparison between Monte Carlo Generators
- Data Validation
- GoF

When and if these techniques make it to real analyses, I suspect we will find plenty of wrong bck estimates ...

But maybe we will find New Physics as well !!

Model-Independent search algorithms also good for:

- Comparison between Monte Carlo Generators
- Data Validation
- GoF

When and if these techniques make it to real analyses, I suspect we will find plenty of wrong bck estimates ...

But maybe we will find New Physics as well !!

Thank You