How Good is the Standard Model?

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

Goodness of Fit

The major concern of any scientist:

Am I doing **everything right?**

Goodness of Fit

The major concern of any scientist:

Am I doing **everything right?**

Being unable to answer, we turn to an easier question:

What could be **wrong?**
and we check **that**

Goodness of Fit

The major concern of any scientist:

Am I doing **everything right?**

Being unable to answer, we turn to an easier question:

What could be **wrong?**
and we check **that**

Cross-checks are more easy the more specifically we characterise the possible failure. But also less powerful

easy/partial

- did I turn QED showering on, in my PYTHIA simulation?
- is the power plug of my detector connected?
- …
-
-
-

hard/complete

Goodness of Fit

The major concern of any scientist:

Am I doing **everything right?**

Being unable to answer, we turn to an easier question:

What could be **wrong?**
and we check **that**

Cross-checks are more easy the more specifically we characterise the possible failure. But also less powerful

easy/partial

- did I turn QED showering on, in my PYTHIA simulation?
- is the power plug of my detector connected?
- …
- is my detector system working “normally”?
- …
-

hard/complete

Goodness of Fit

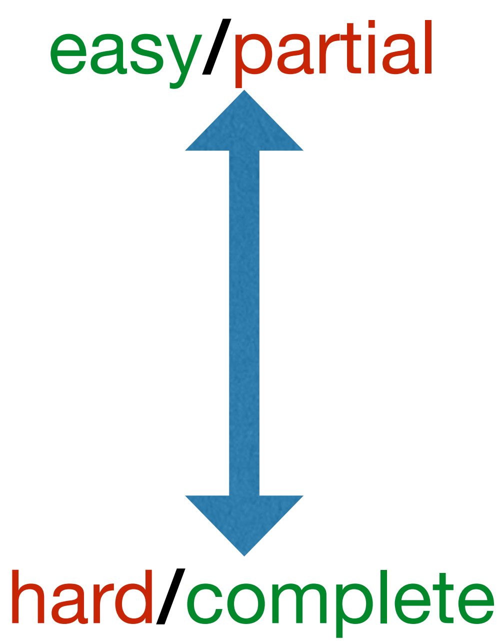
The major concern of any scientist:

Am I doing **everything right?**

Being unable to answer, we turn to an easier question:

What could be **wrong?**
and we check **that**

Cross-checks are more easy the more specifically we characterise the possible failure. But also less powerful



- did I turn QED showering on, in my PYTHIA simulation?
- is the power plug of my detector connected?
- …
- is my detector system working "normally"?
- …
- is my state-of-the-art knowledge of fundamental interactions (the **SM**) **correct**, or it **fails** to describe the LHC data?

Goodness of Fit

Statisticians formulate the problem as **g.o.f.***

Be $\mathcal{D}$ a set of data, and R a stat. hyp. for their distribution

Does R provide the **right description** of $\mathcal{D}$?

*often question emerges after optimising distribution free parameters on the data, as a way to assess fit quality. But the problem is more general

Goodness of Fit

Statisticians formulate the problem as **g.o.f.**

Be $\mathcal{D}$ a set of data, and R a stat. hyp. for their distribution

Does R provide the **right description** of $\mathcal{D}$?

Again, answer more easy the more restrictive assumptions we make on how the true distribution, if not R, can look like

But, more partial as well.

Goodness of Fit

Statisticians formulate the problem as **g.o.f.**

Be $\mathcal{D}$ a set of data, and R a stat. hyp. for their distribution

Does R provide the **right description** of $\mathcal{D}$?

Again, answer more easy the more restrictive assumptions we make on how the true distribution, if not R, can look like

But, more partial as well.

G.o.f. is a "ill-posed" problem: no **optimal** solution exists.
But plenty of **good** solutions exist, especially (only?) in 1d *

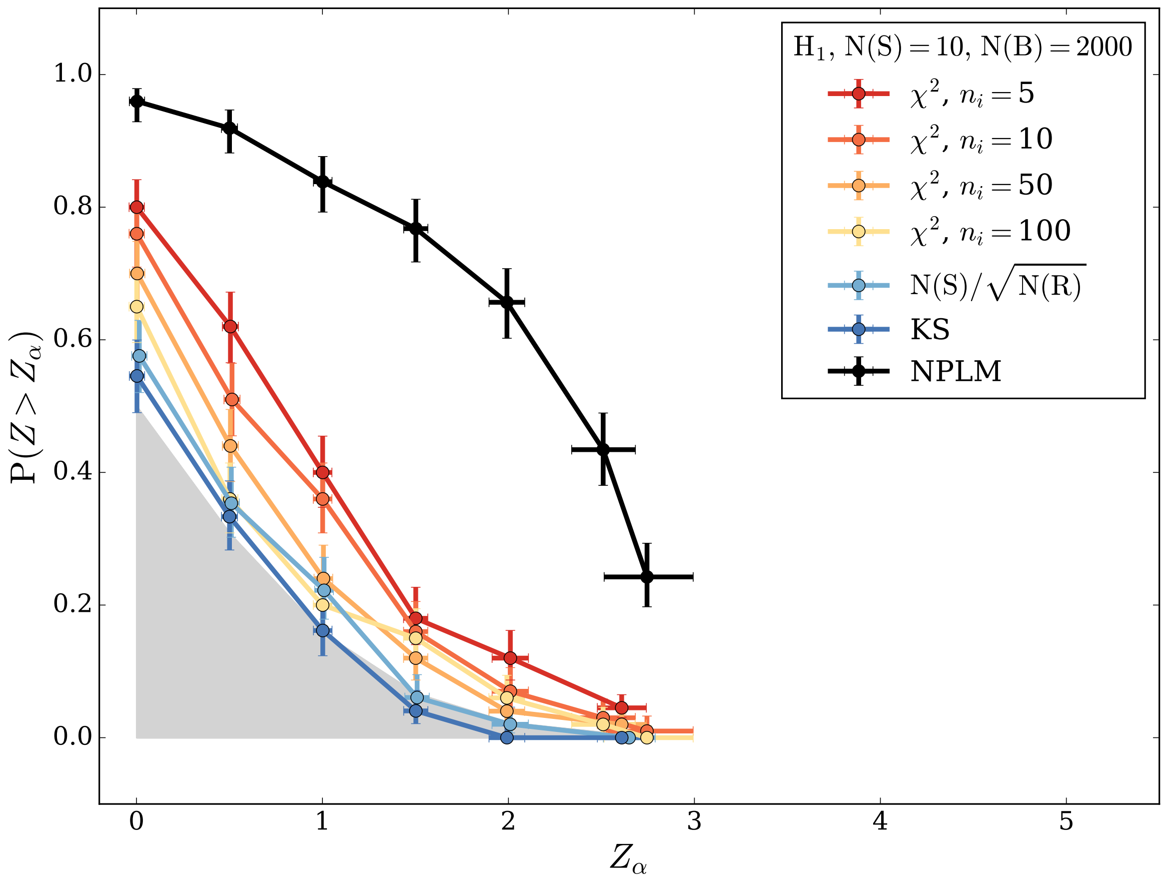
We can search for **better** solutions, perhaps even in 1d

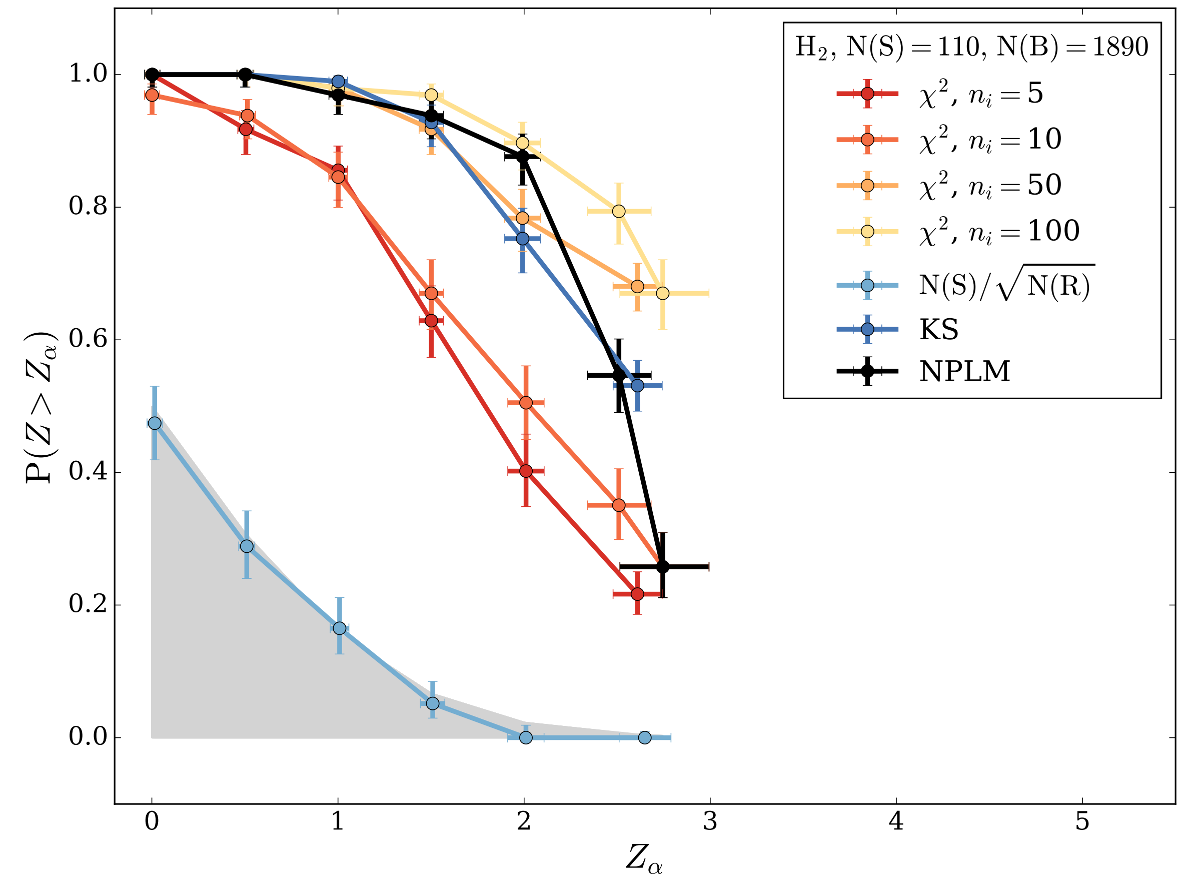
*For instance, students quickly learn to plot **binned histograms** with their data, because this **often** allow them to find mistakes

Goodness of Fit

Probability to find evidence of R being wrong at some level of confidence.

True data distribution departs from R in different ways, in the two plots.





We can search for **better** solutions, perhaps even in 1d

*For instance, students quickly learn to plot **binned histograms** with their data, because this **often** allow them to find mistakes

The LHC g.o.f. challenge

By analysing the LHC data, we would like to find evidence of **failure of the SM theory**, suggesting need of **BSM.**

The LHC g.o.f. challenge

By analysing the LHC data, we would like to find evidence of **failure of the SM theory**, suggesting need of **BSM.**

But this is a tremendously hard task!

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

The LHC g.o.f. challenge

By analysing the LHC data, we would like to find evidence of **failure of the SM theory**, suggesting need of **BSM.**

But this is a tremendously hard task!

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

We thus address the much easier task of searching, one by one, **specific BSM models** of "new physics”

Choose observables sensitive to **one BSM model**
This observable in general **not** sensitive to **another BSM model**

The LHC g.o.f. challenge

By analysing the LHC data, we would like to find evidence of **failure of the SM theory**, suggesting need of **BSM.**

But this is a tremendously hard task!

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

We thus address the much easier task of searching, one by one, **specific BSM models** of "new physics"

Choose observables sensitive to **one BSM model**
This observable in general **not** sensitive to **another BSM model**

We call this strategy a "**Model-Dependent**" search

The LHC g.o.f. challenge

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

***very likely**

Most likely, **we will not see the SM fail to describe data**

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

The LHC g.o.f. challenge

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

***very likely**

Most likely, **we will not see the SM fail to describe data**

We must design **Model Independent** searches

aimed at detecting "generic" data departures from SM

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

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

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

G.o.f. from Maximum Likelihood

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

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

G.o.f. from Maximum Likelihood

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

Reference Distribution: $n(x|\mathrm{R})$

Alternative Distribution: $n(x|\mathbf{w})$

depending on **parameters** (composite)

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

G.o.f. from Maximum Likelihood

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

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

Reference Distribution: $n(x|\mathrm{R})$

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

Test statistic:

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

Model Dependent Strategy

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

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

G.o.f. from Maximum Likelihood

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

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

Reference Distribution: $n(x|\mathrm{R})$

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

Test statistic:

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

Model Dependent Strategy

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

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

This is what regularly done at the LHC: target sensitivity to manifestations of one specific theoretical **New Physics Model**

G.o.f. from Maximum Likelihood

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

Reference Distribution: $n(x|\mathrm{R})$

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

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

Test statistic:

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

Model Dependent Strategy

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

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

This is what regularly done at the LHC: target sensitivity to manifestations of one specific theoretical **New Physics Model**

Model Independent Strategy

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

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

G.o.f. from Maximum Likelihood

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

Reference Distribution: $n(x|\mathrm{R})$

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

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

Test statistic:

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

Model Dependent Strategy

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

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

This is what regularly done at the LHC: target sensitivity to manifestations of one specific theoretical **New Physics Model**

Model Independent Strategy

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

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

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

↓

Binned Histogram Test
(AKA, Baker-Cousins test)
(used by ATLAS and CMS for Model-Independent New Physics Searches)

G.o.f. from Maximum Likelihood

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

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

Reference Distribution: $n(x|\mathrm{R})$

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

Test statistic:

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

Model Dependent Strategy

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

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

This is what regularly done at the LHC: target sensitivity to manifestations of one specific theoretical **New Physics Model**

Model Independent Strategy

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

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

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

↓

Our Proposal

G.o.f. from Maximum Likelihood

Data: $\mathcal{D} = \{x_i\},\ i = 1, \ldots, N_{\mathcal{D}}$

$N\,P(x)$

$dx\, n(x)$

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

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
- Using other models also possible

Model Dependent Strategy

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

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

This is what regularly done at the LHC: target sensitivity to manifestations of one specific theoretical **New Physics Model**

Model Independent Strategy

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

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

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

Our Proposal

Maximum Likelihood Loss

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

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

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

We need a **Reference Sample**, distributed according to Reference Model

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

Approximate integral as Monte Carlo sum:

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

Maximum Likelihood Loss

Turn the evaluation of “t” into supervised training problem:

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

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

We need a **Reference Sample**, distributed according to Reference Model

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

Approximate integral as Monte Carlo sum:

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

In order to read this as “equal”, we need

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

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

Maximum Likelihood Loss

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

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

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

We need a **Reference Sample**, distributed according to Reference Model

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

Approximate integral as Monte Carlo sum:

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

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

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

$$L[f] = \sum_{(x,y)}\left[(1-y)\frac{N(\mathrm{R})}{\mathcal{N}_{\mathcal{R}}}(e^{f(x)} - 1) - y\, f(x)\right]$$

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

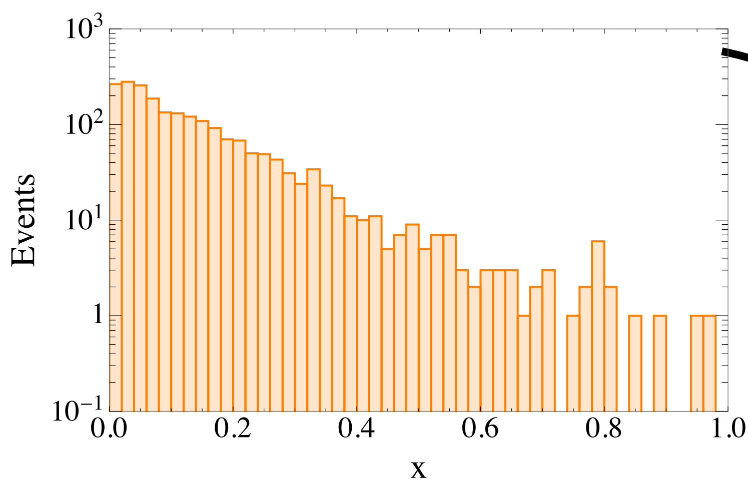
  Preferably, more abundant than the data: $\mathcal{N}_{\mathcal{R}} \gg N(\mathrm{R})$

*It must be SM-distributed **if the SM is true**. If BSM in Reference Sample, this generically does not harm our ability to see tension.

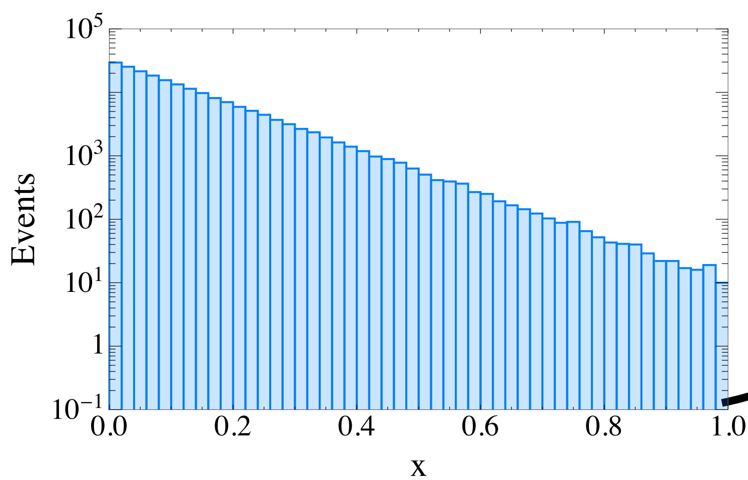
The Algorithm

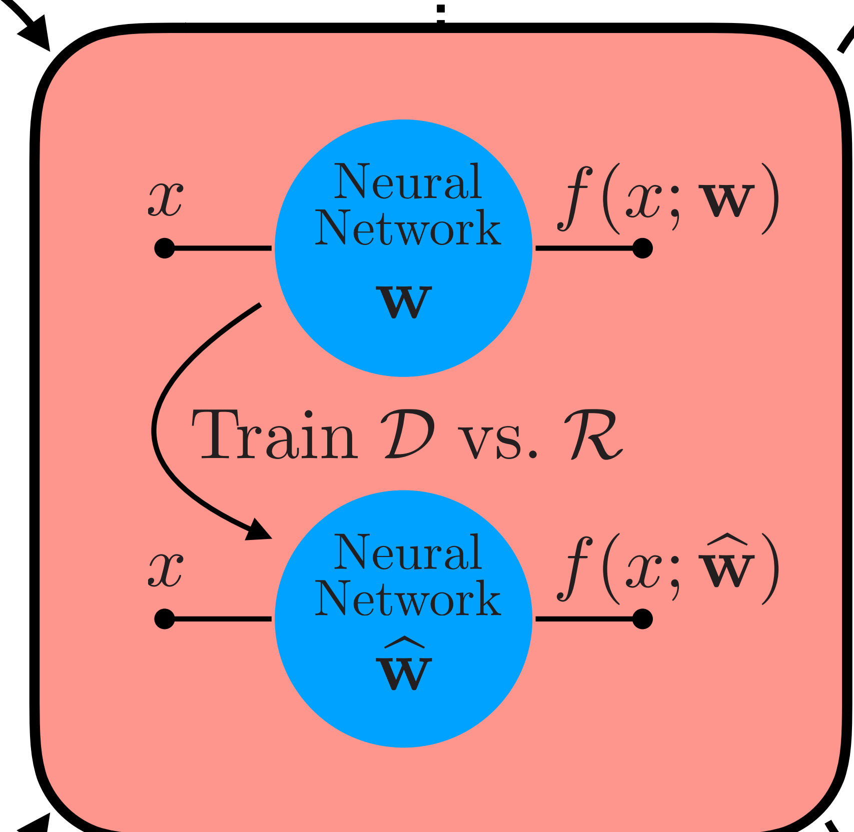
INPUT

Data sample $\mathcal{D}$



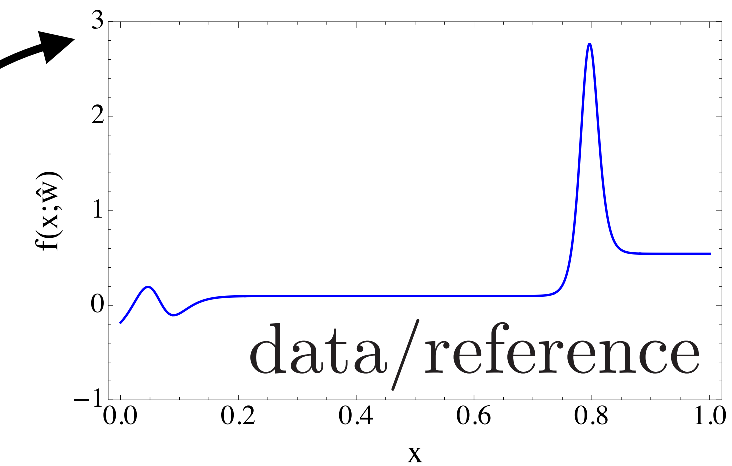
Reference sample $\mathcal{R}$





OUTPUT

Dist. log ratio



$$f(x;\widehat{\mathbf{w}}) \simeq \log\left[\frac{n(x|\mathrm{T})}{n(x|\mathrm{R})}\right]$$

Test statistic t
computed on the data sample $\mathcal{D}$

$$t(\mathcal{D}) = -2\,\underset{\{\mathbf{w}\}}{\mathrm{Min}}\,L[f]$$

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: $\mathcal{N}_{\mathcal{R}} \gg N(\mathrm{R})$

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

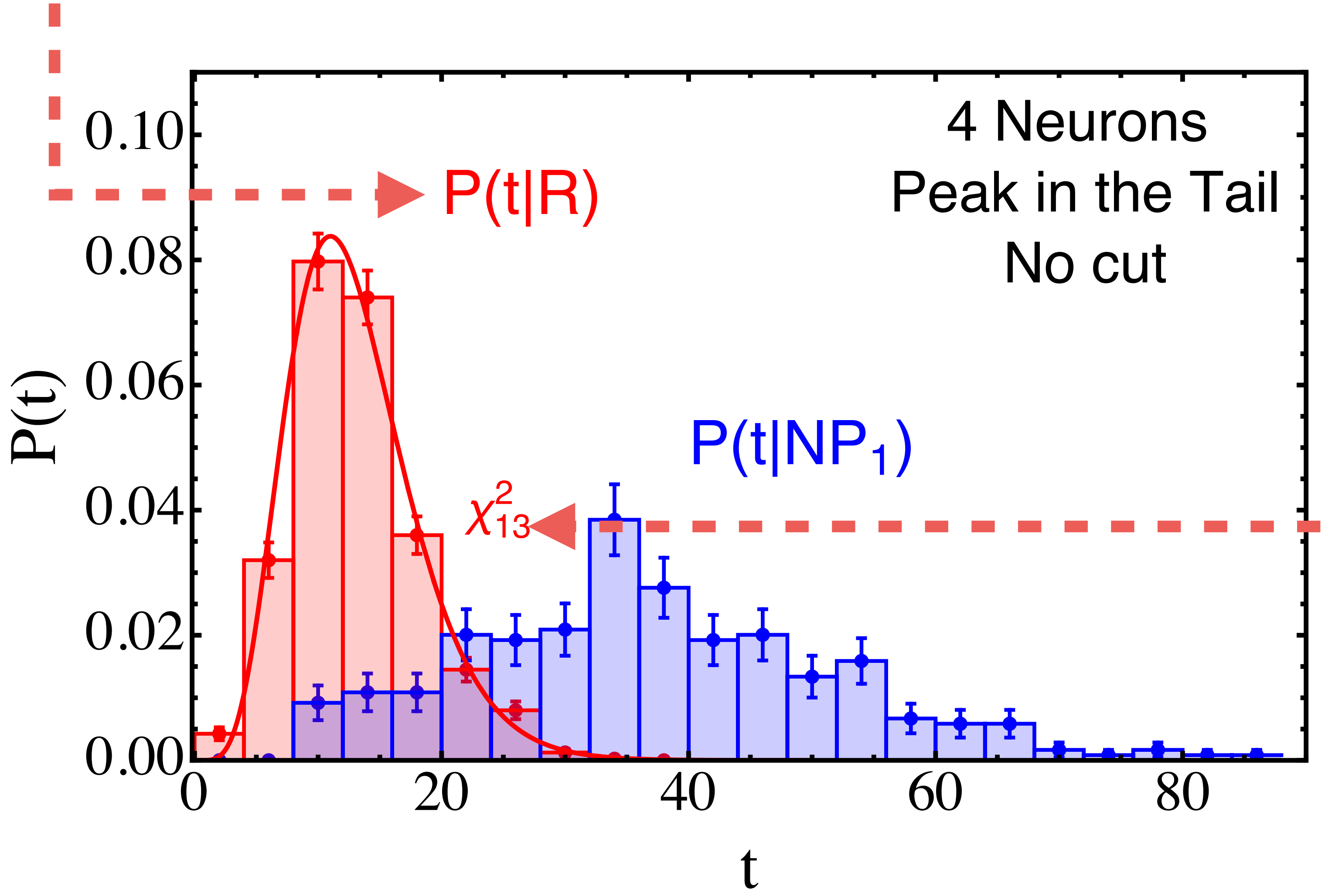
This will give us the observed p-value:

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

Illustrating Performances

(Simple 1d example with exponential Reference)

Distribution of the test statistic "t" in Reference Hypothesis

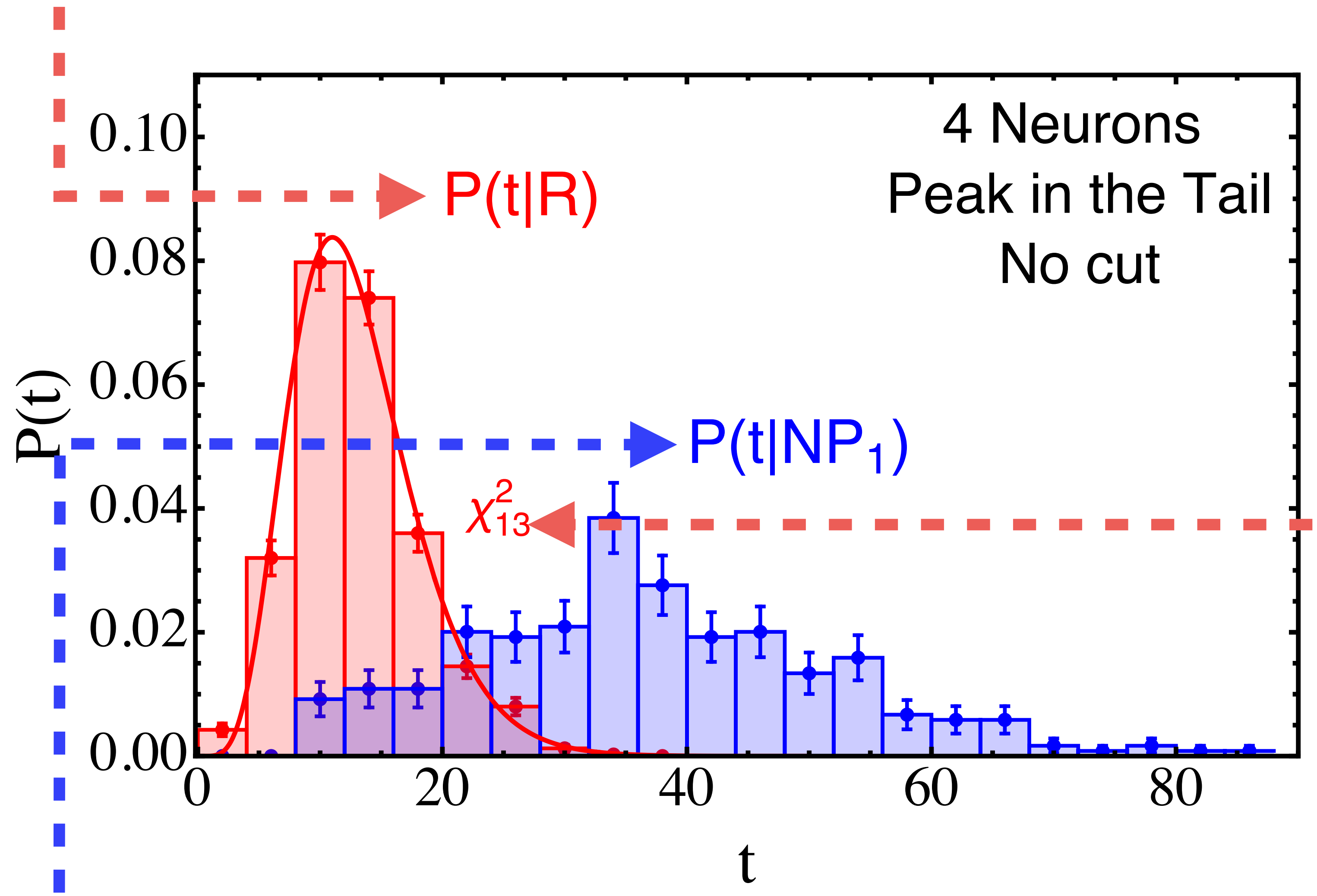


Notice agreement with **"Asymptotic Formula"**
(more on this later)

Illustrating Performances

(Simple 1d example with exponential Reference)

Distribution of the test statistic "t" in Reference Hypothesis



Notice agreement with **"Asymptotic Formula"**

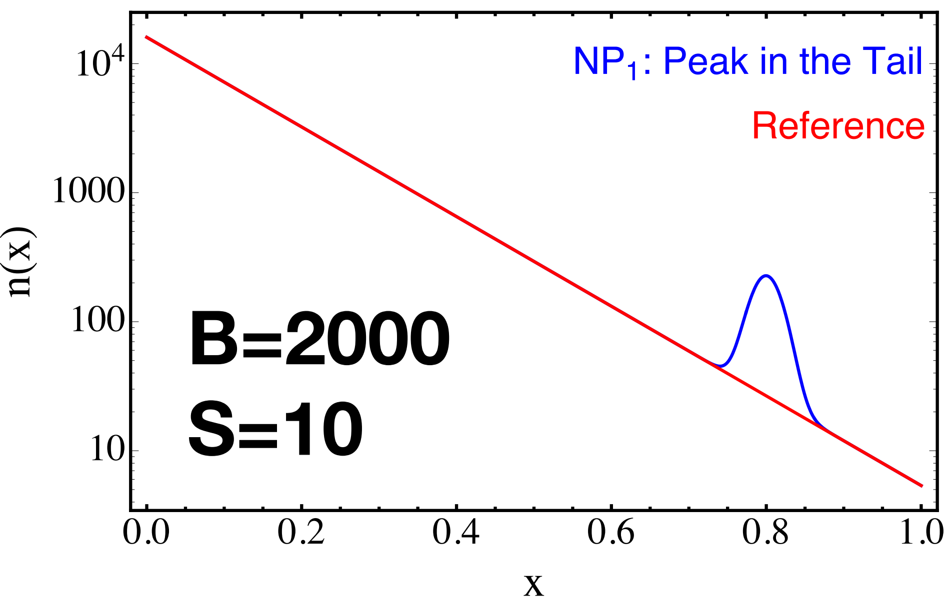
(more on this later)

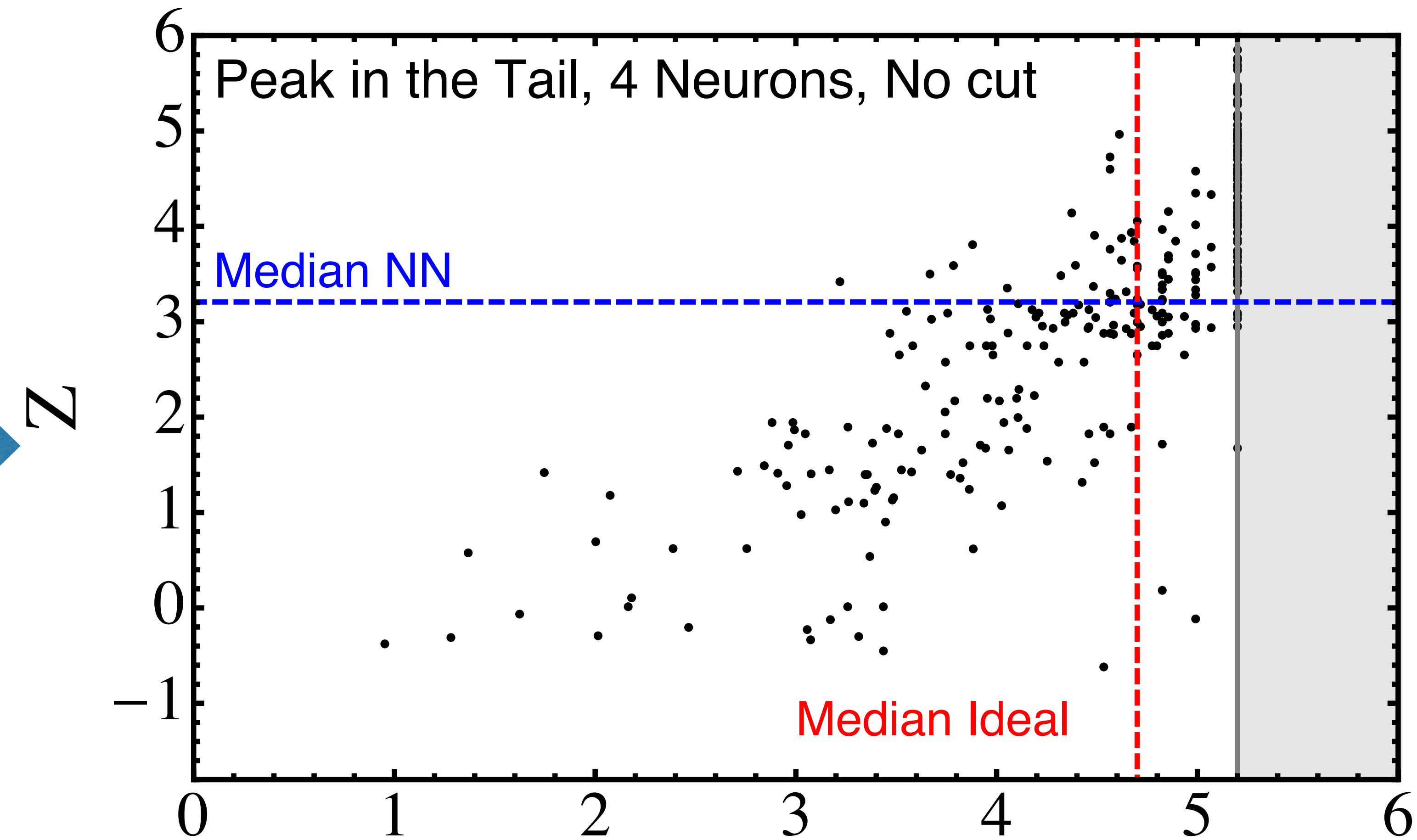
Distribution of "t" in one New Physics Model Hypothesis

t → p → Z-score (we use $Z = \Phi^{-1}(1-p)$)

Quantifying Performances

(Simple 1d example with exponential Reference)





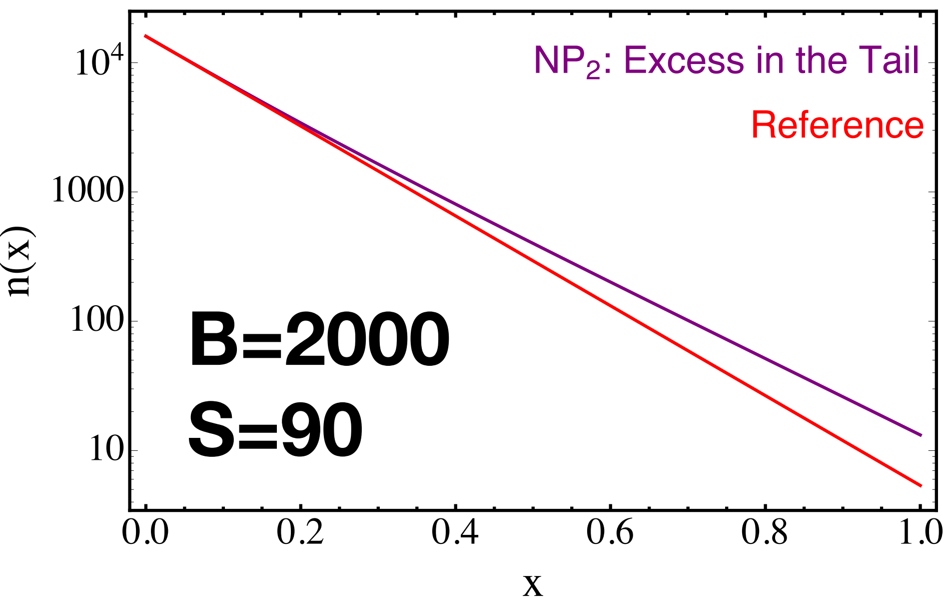
“Ideal Z-score”: $Z_{\rm id}$

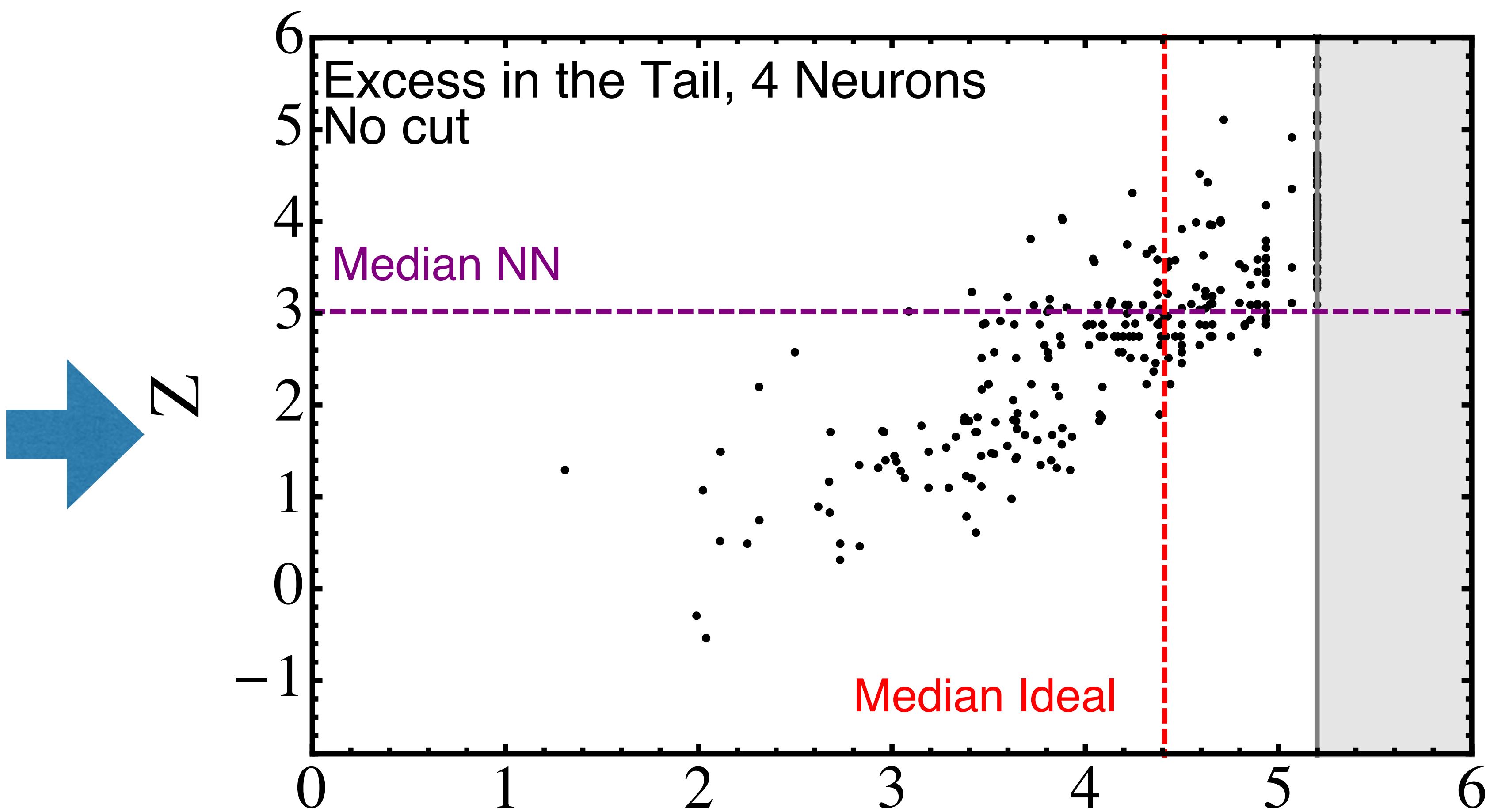
A “measure of dataset discrepancy”

(the Z-score of optimal test for NP1 model)

Quantifying Performances

(Simple 1d example with exponential Reference)





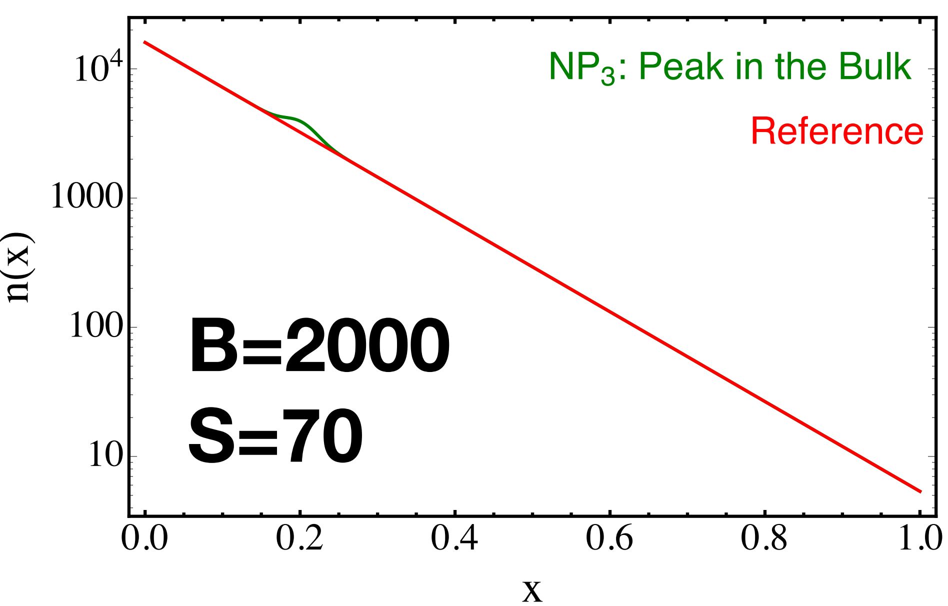
“Ideal Z-score”: $Z_{\rm id}$

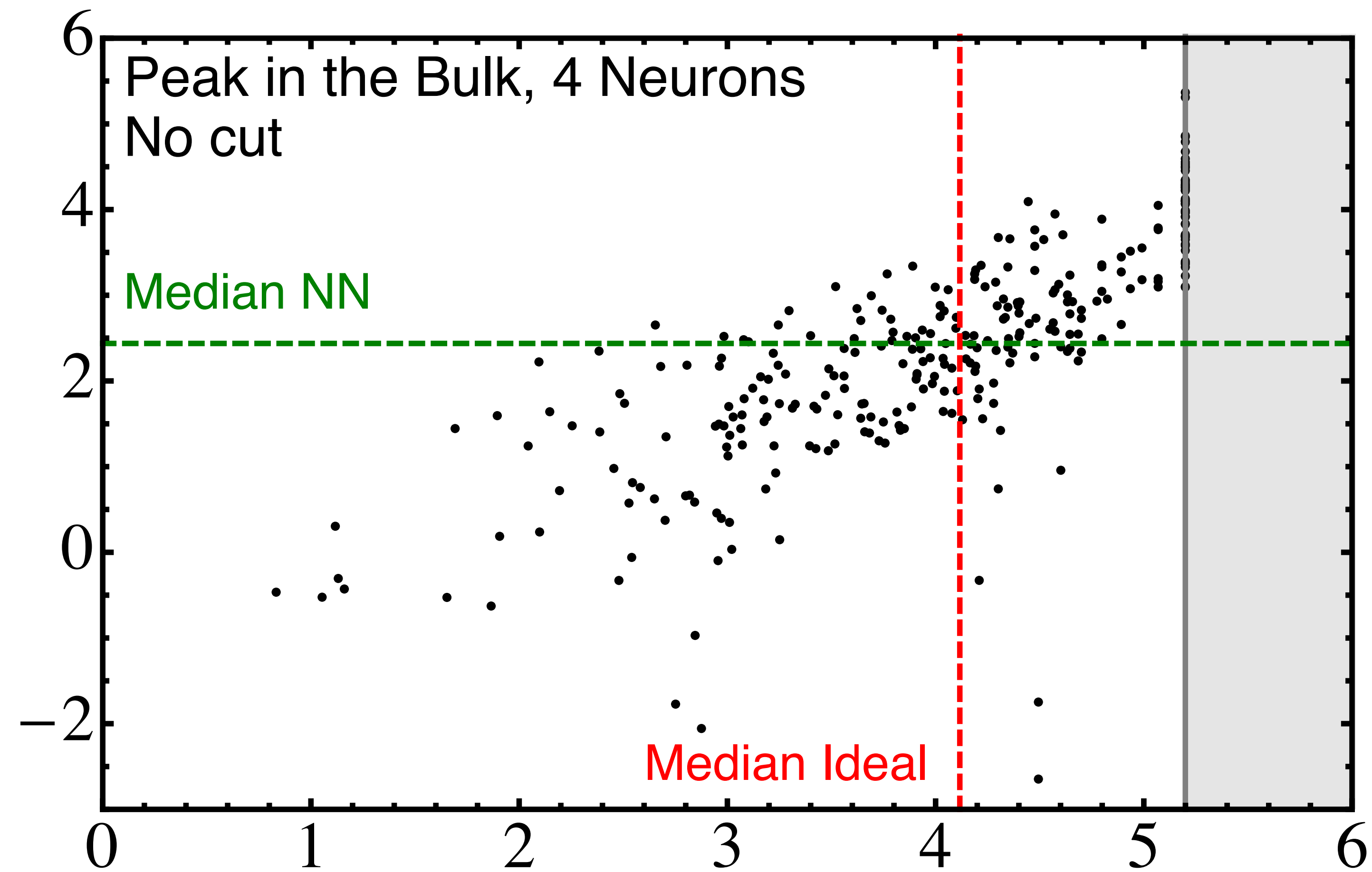
A “measure of dataset discrepancy”

(the Z-score of optimal test for NP2 model)

Quantifying Performances

(Simple 1d example with exponential Reference)





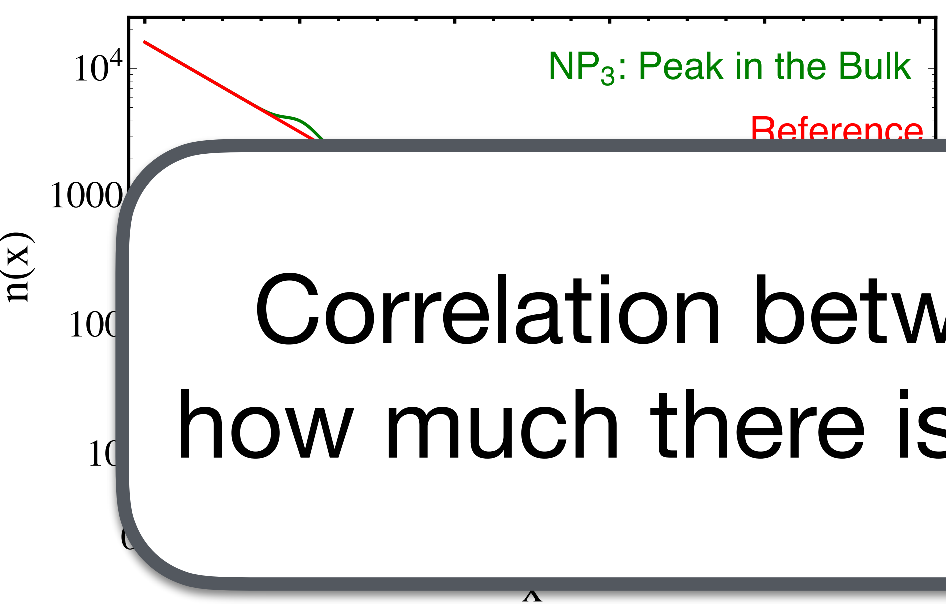
"Ideal Z-score": Z_{id}

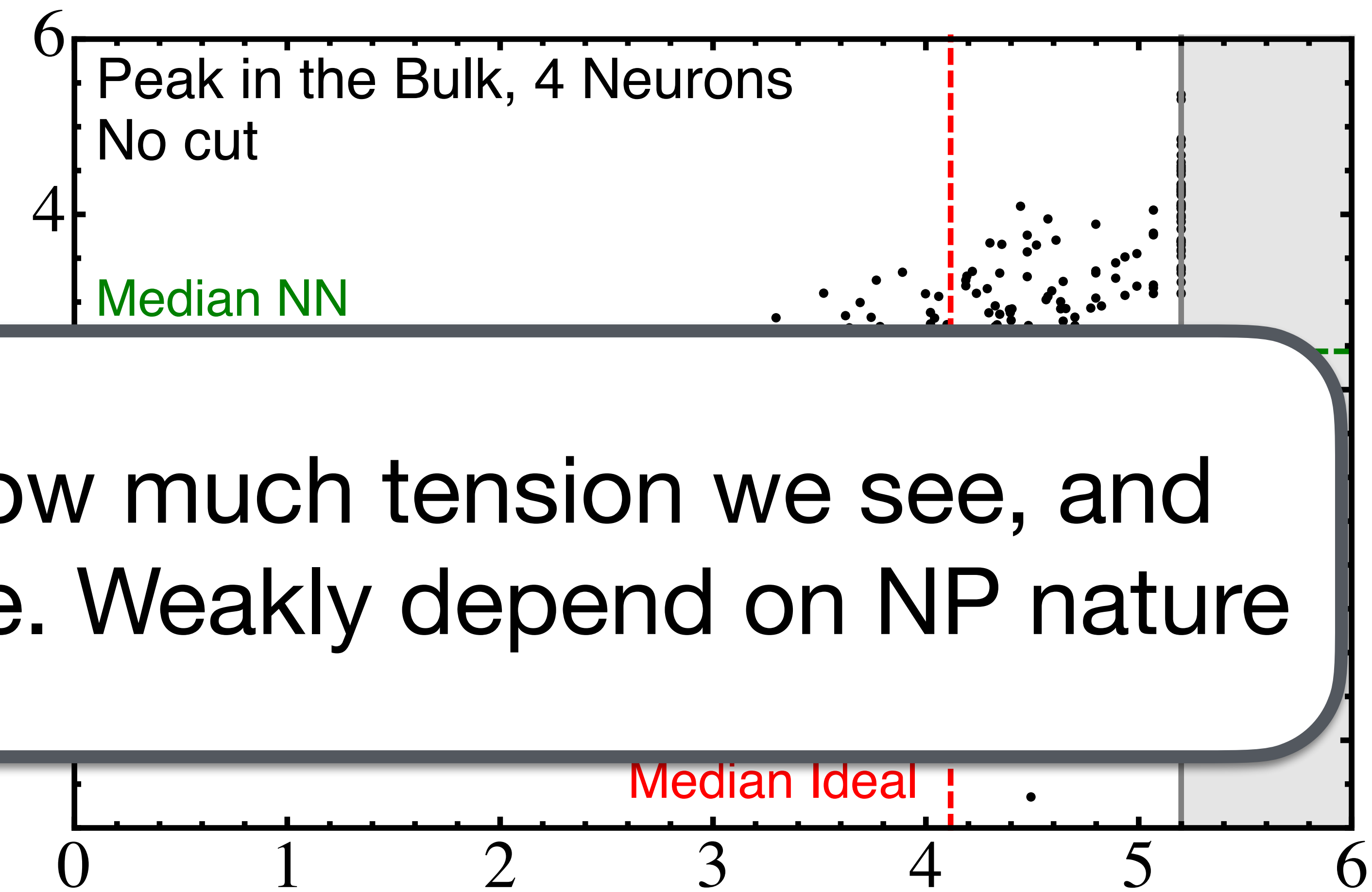
A "measure of dataset discrepancy"

(the Z-score of optimal test for NP3 model)

Quantifying Performances

(Simple 1d example with exponential Reference)





Correlation between how much tension we see, and how much there is to see. Weakly depend on NP nature

“Ideal Z-score”: $Z_{\rm id}$

A “measure of dataset discrepancy”

(the Z-score of optimal test for NP3 model)

An **Imperfect** Machine

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

An **Imperfect** Machine

Reference Model Predictions are unavoidably imperfect

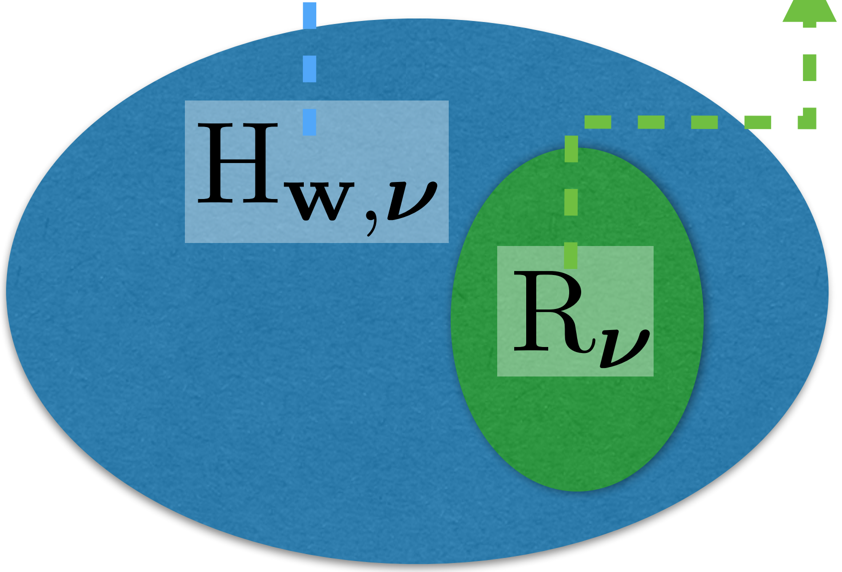
e.g., PDF/Lumi/Detector Modeling …

Imperfections are **Nuisance Parameters**

Constrained by **Auxiliary Measurements**

Define a **composite** Reference hypothesis

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



Just like in no-nuisance case:

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

Beyond-Reference effects
parametrised by NN

An **Imperfect** Machine

Reference Model Predictions are unavoidably imperfect

e.g., PDF/Lumi/Detector Modeling …

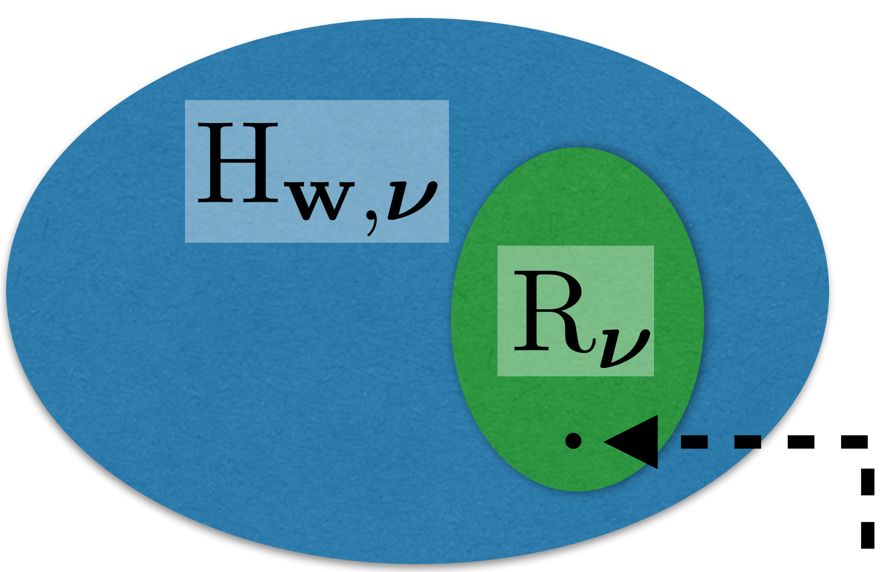
Imperfections are **Nuisance Parameters**

Constrained by **Auxiliary Measurements**

Define a **composite** Reference hypothesis

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

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



Central-Value Reference: $\mathrm{R}_{\mathbf{0}}$

Nuisance set to their C-V

An **Imperfect** Machine

Reference

Imperfect

$t(\mathcal{D},\mathcal{A}) = 2 \ \mathrm{m}$
$\mathbf{w}$

"Delta" term by direct likelihood maximisation

After **learning the effect of nuisance** locally on distribution

$$r(x;\boldsymbol{\nu}) \equiv \frac{n(x|\mathrm{R}_{\boldsymbol{\nu}})}{n(x|\mathrm{R}_{\mathbf{0}})} = \exp\left[\nu\,\delta_1(x) + \frac{1}{2}\nu^2\,\delta_2(x) + \ldots\right]$$

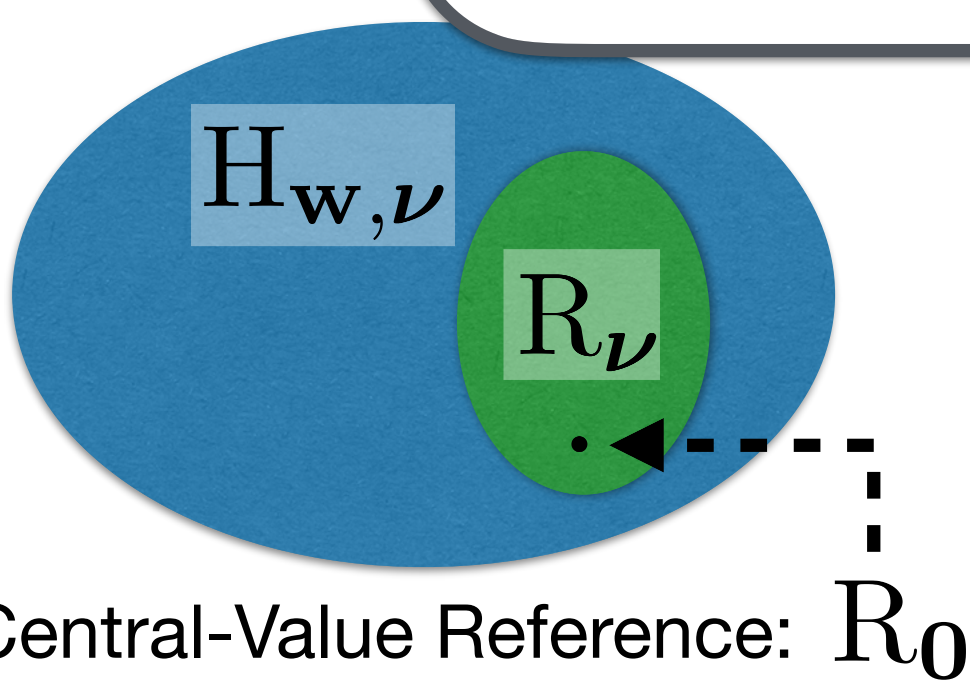
Adaptation of **likelihood-free inference** techniques

Would require dedicated seminar. [See e.g. 1907.10621, 2007.10356, ...]

Just be aware that:

i) learning requires (enough) R data with non-C-V nuisance

ii) the **quality** of the reconstruction can play crucial role



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

Central-Value Reference:

Nuisance set to their C-V

An **Imperfect** Machine

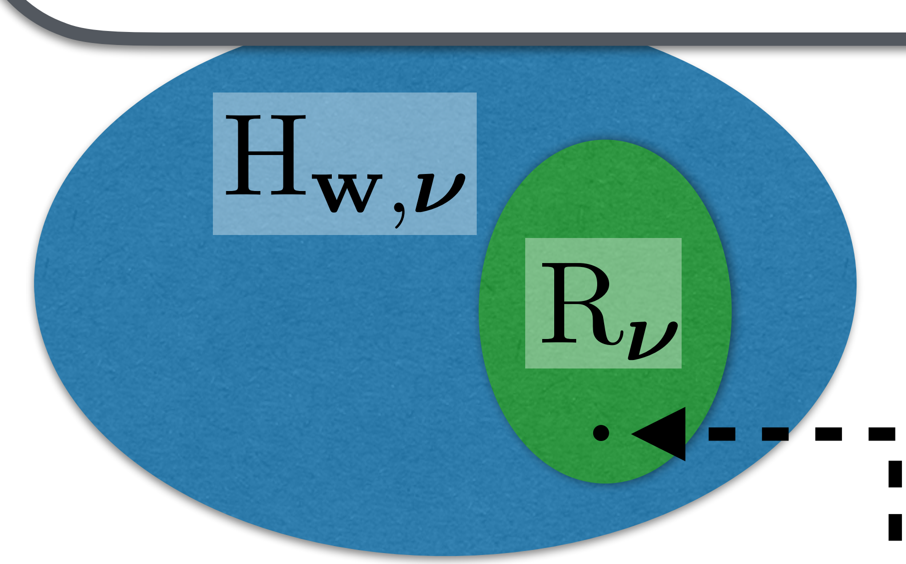
“Tau” term by training on Data

Almost like for no nuisance, but with modified ML-Loss:

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

And, with simultaneous **training over the nuisance parameters**

Data trained against **Central-Value Reference** sample **only**



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

Central-Value Reference: $\mathrm{R}_{\mathbf{0}}$

Nuisance set to their C-V

An **Imperfect** Machine

Reference Model Predictions are unavoidably imperfect

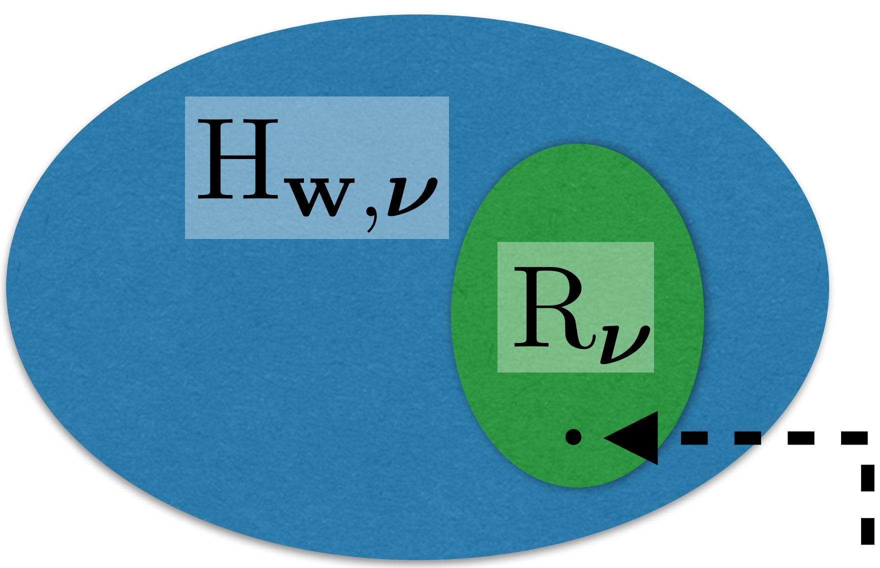
e.g., PDF/Lumi/Detector Modeling …

Imperfections are **Nuisance Parameters**

Constrained by **Auxiliary Measurements**

Define a **composite** Reference hypothesis

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



Central-Value Reference: $\mathrm{R}_{\mathbf{0}}$

Nuisance set to their C-V

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

If we do all right, by Wilks-Wald we get:

$$P(t|\mathrm{R}_{\boldsymbol{\nu}}) = P(t|\mathrm{R}_{\mathbf{0}}) = \chi^2_d$$

An **Imperfect** Machine

Reference Model Predictions are unavoidably imperfect

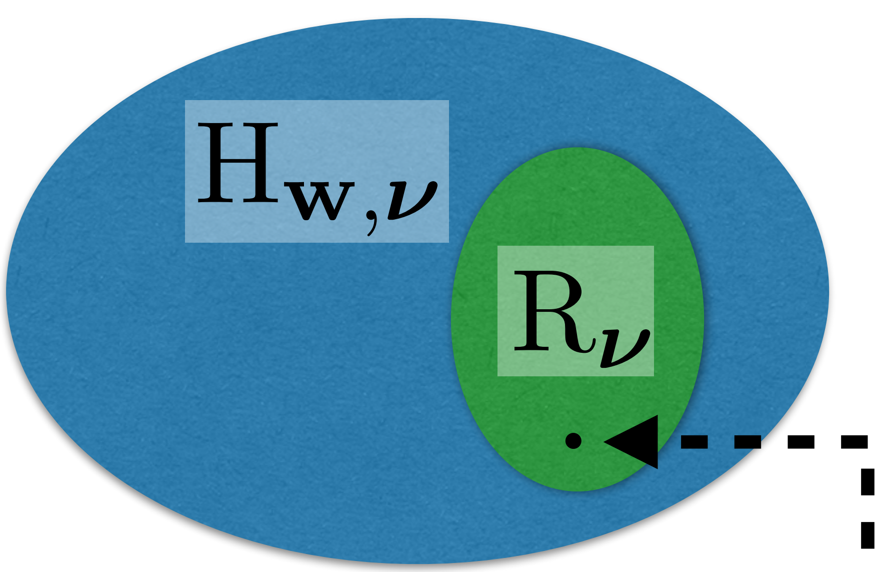
e.g., PDF/Lumi/Detector Modeling …

Imperfections are **Nuisance Parameters**

Constrained by **Auxiliary Measurements**

Define a **composite** Reference hypothesis

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



Central-Value Reference: $\mathrm{R}_{\mathbf{0}}$

Nuisance set to their C-V

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

If we do all right, by Wilks-Wald we get:

$$P(t|\mathrm{R}_{\boldsymbol{\nu}}) = P(t|\mathrm{R}_{\mathbf{0}}) = \chi^2_d$$

Independence of t distribution on the **true value** of **nuisance** is **essential** for feasible test

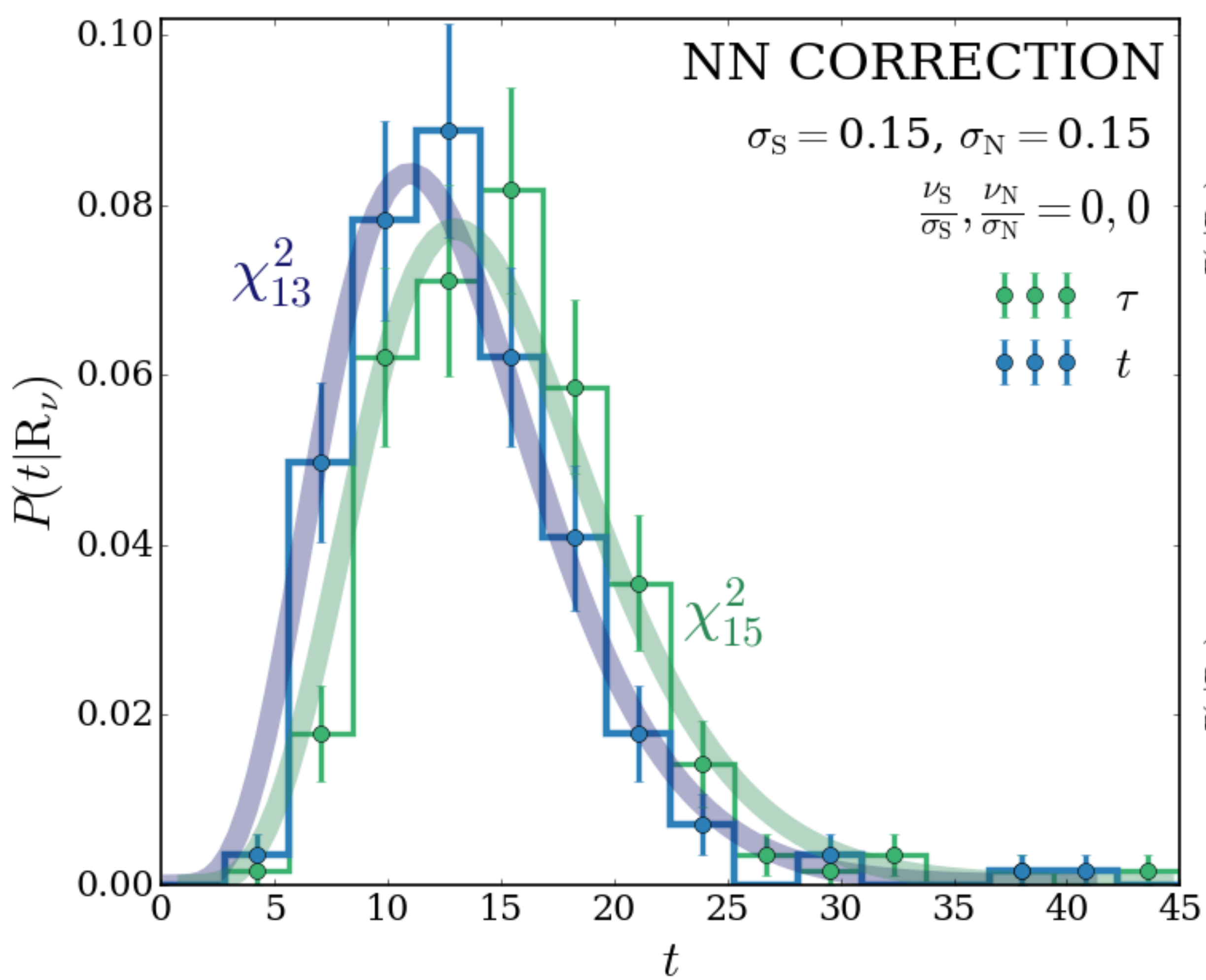
An **Imperfect** Machine at Work

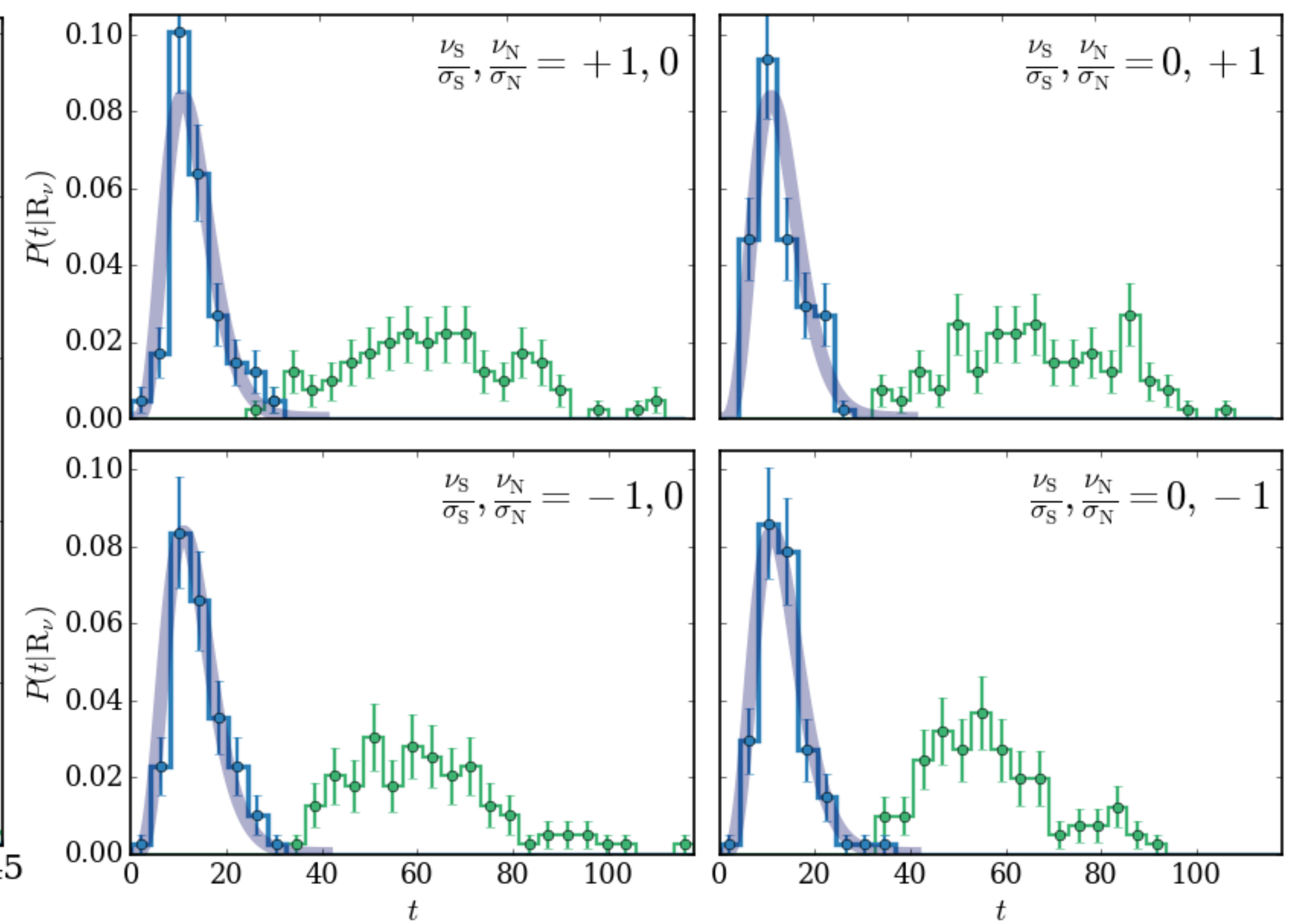
(Simple 1d example with exponential Reference)

Tau distribution distorted by non-central value nuisance

if not corrected, produces false positives

t = Tau-Delta independent of nuisance





Remarks/Concerns

Remark #1: By Wilks-Wald Theorem, P(t|R) is a χ^2, 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 **χ^2-compatibility** as **Model Selection criterion**

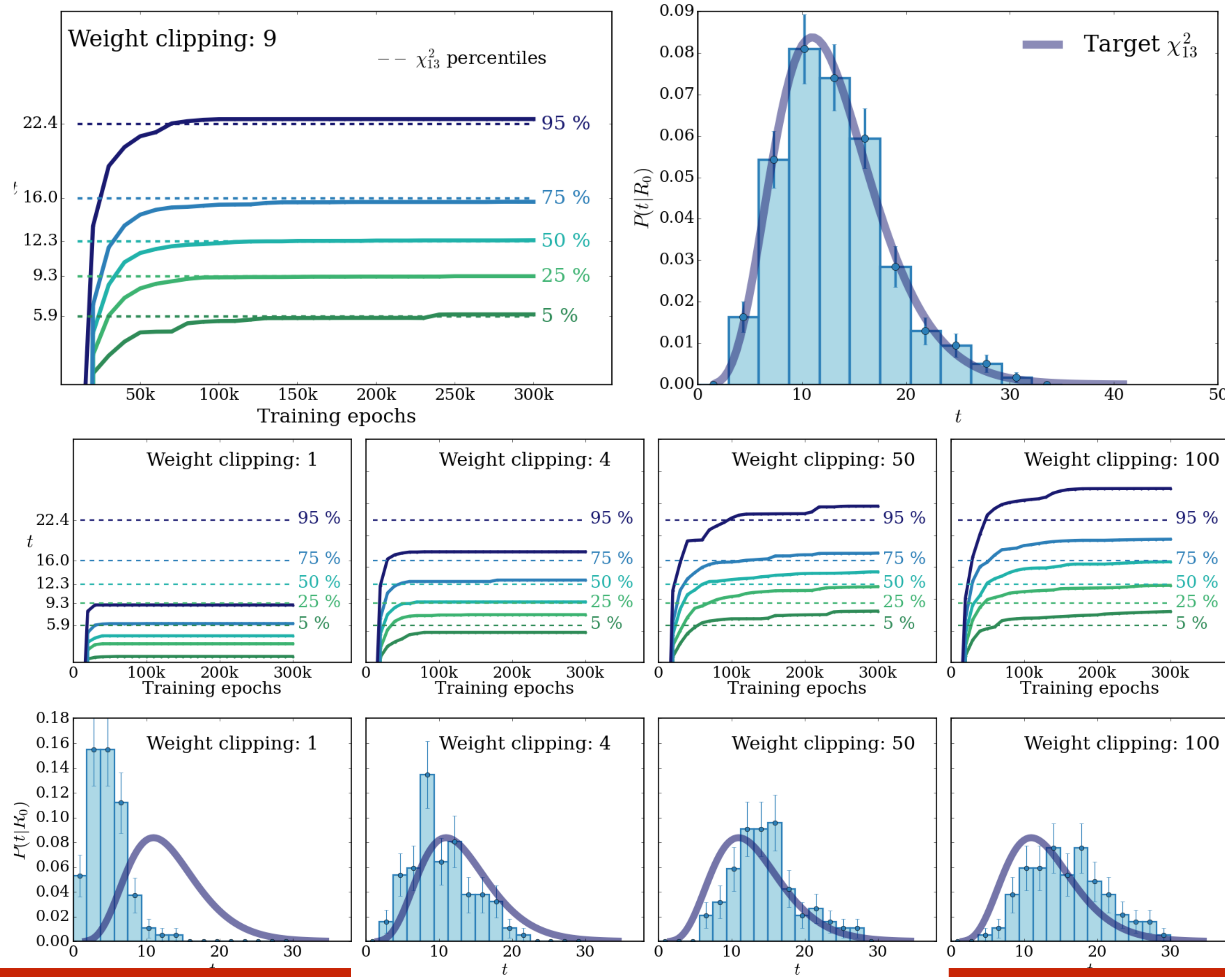
Asy.For. violation = sensitivity to low-statistics portion of dataset = overfitting

Selection w/o nuisance ensures nuisance-independent chi-sq

Criterion used in particular to select **Weight Clipping** regularisation par.

Weight Clipping Selection

(Simple 1d example with exponential Reference)



Asy.For. violation by fit parameters boundary

Asy.For. violation by sensitivity to sparse data points

Remarks/Concerns

Remark #1: By Wilks-Wald Theorem, P(t|R) is a χ^2, 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 **χ^2-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

Remarks/Concerns

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

Remarks/Concerns

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.

Remarks/Concerns

Remark #3:

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

[proposed by J.Friedman, 2004 in hep context, but not really studied]

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

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

Remarks/Concerns

Remark #3:

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

[proposed by J.Friedman, 2004 in hep context, but not really studied]

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 comparative study useful.

[see Chakravarti, Kuusela, Lei, Wasserman, 2021 for a first attempt]

Outlook

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?)

Outlook

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

Outlook

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!

Outlook

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

Outlook

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

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

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

Outlook

Model-Independent search algorithms also good for:

- Comparison between Monte Carlo Generators
- Data Validation
- GoF

Outlook

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 …

Outlook

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 !!

Outlook

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